

Citation for published version:

Vass, P 2015, *Political Economy in Practice: The Foundations for Policy, Management and Independent Economic Regulation*. Self Published.

Publication date:

2015

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Political Economy in Practice - The Foundations for Policy, Management and Independent Economic Regulation

Peter Vass

**A course text in political economy - economics and finance, financial and
management accounting, regulation and governance**

**Author's copyright
Summer 2015**

***Note: The diagrams are illustrative to show optimum points and important cross-over points.
They are not an accurate representation of non-linear lines.***

Preface

Political Economy in Practice

Policy, Management and Independent Economic Regulation

Purpose of the course text

The text is designed for courses in the areas of applied political economy. Its five sections are intended to integrate economics with accountancy, and provide a foundation of analytical examples. The application of political economy ideas to important policy questions follows, and concludes with the design of the regulatory framework for a modern state. By combining a wide range of the elements which feature in political economy as a subject, the text can provide a useful complement to a specialist course textbook which focuses on one element alone, such as Economics 12th edition, Lipsey and Chrystal, OUP, 2011. It can be used as a main course text in its own right, and is intended for both undergraduate and postgraduate courses. The emphasis is on practical applications, which makes the course text suitable for general courses in universities, as well as in other centres of education, particularly government and professional bodies. Worked examples and algebraic formulae are used where necessary. The analytical examples in section III have been chosen to cover those areas which are often less fully covered in standard textbooks. These include:

- * Reconciling investment decision rules in terms of NPV with ‘modified’ internal rates of return.
- * Introducing the ‘opportunity cost of deferral’ as the appropriate measure for investment decisions on mutually exclusive projects subject to capital rationing.
- * Integrating joint processes and by-products, further processing and recovery of ‘common’ fixed costs in optimising production decisions and pricing.
- * Identifying ‘annuity’ depreciation profiles as the economic foundation of depreciation.

The text is designed to be suitable for courses in different countries because where a UK example of applied political economy has been used it is because it has international relevance. In particular, this includes the use of price control rather than profit control in regulating privatised utilities and network industries, as well as the economic history of privatisation from 1984 to 2013; the development of accounting standards from professional self-regulation to the adoption of international accounting standards; and the recognition that an effective regulatory state requires effective accountability along with appropriate institutional structures.

The Author: Peter Vass graduated with a BSc(econ) and an MSc(econ) in urban and regional planning studies from University College London (UCL). Joining the Government Economic Service, first at the Ministry of Transport, and then at the Department of the Environment, his work focused on the cost-benefit analysis of transport schemes, and then local government finance, particular the ‘rate support grant’. He moved to Essex County Council to qualify as a public finance accountant (CPFA), and then joined the secretariat of the Chartered Institute of Public Finance and Accountancy (Cipfa). He specialised in accounting practice and accounting standards, being seconded part time to the Accounting Standards Committee on developing public sector accounting standards. Becoming head of Cipfa’s technical and policy division, an important focus of his work was on finance and statistics in utility and network industries, particularly water, energy and transport, which gave him an opportunity to lecture at the London Business School. In 1990, he helped establish the Centre for the study of Regulated Industries (CRI) following the privatisation of utility and network industries, and was appointed research director when he moved to the University of Bath School of Management in 1991 as a senior lecturer in accounting and finance. The CRI was transferred to the University of Bath in 1997, and he directed it until 2010. The publications of the CRI over twenty years can be found at www.bath.ac.uk/cri. He has been a special adviser to select committees of the UK’s House of Commons and the House of Lords.

Political Economy in Practice - The Foundations for Policy, Management and Independent Economic Regulation	
I Political economy	4
1. Introduction – The Scope of Political Economy	4
2. The Wealth of Nations – Adam Smith 1776	5
3. The Regulatory State: Ensuring its Accountability – House of Lords 2004	7
II The Foundations	11
Economic analysis	
1. Present and future values; discount rates, the IRR and annuities	12
2. The standard investment rule	16
3. Demand and supply/cost and revenue curves; profit maximising production	17
4. Competition, choice and monopoly; the three costs of business	21
5. Cost-benefit analysis, welfare and the regulatory state	28
Applied economics - Controlling climate change	32
1. ‘Macro’ political economy	
2. An illustrative global approach	33
3. The political reality of action on climate change	37
- ‘Micro’ political economy	40
1. Questionable outcomes - the power of ‘safety and security’	
Accountancy - financial reporting, management accounting and finance	44
Classifying accounting information	
2. The ‘double entry’ system	45
3. Main accounts: profit and loss account, balance sheet, cash flow statement	47
4. Accounting – preparing a balance sheet, P&La/c, and cash flow statement	52
5. Accounting for economic profit; historical cost and revaluation, ratios	57
6. Management accounting: product costing, divisional&overhead allocations	59
7. Budgeting	64
8. Corporate finance and the cost of capital	66
Applied accountancy - management control and performance measurement	72
1. Variance analysis	
2. Control by accounting standards, regulatory accounts and auditing	83
III Analytical examples	91
1. Reconciling NPV and IRR decision rules; window on investment appraisal	91
2. Decision rules with capital rationing and mutually exclusive investments	98
3. Optimising with multiple constraints; dual solutions and opportunity costs	103
4. Economic allocation of joint costs	108
5. Fixed common cost allocation using ‘Ramsey’ pricing	112
6. Economic depreciation and expected technical change	121
IV Independent Economic regulation - Public services privately provided	
1. Why regulate? 128 ; 2. Nationalisation and privatisation 129 ;	
3. Types of competition 131 ; 4. Incentive regulation; the example of price control through RPI-x 132 ; 5. The littlechild model 133 ; 6. Economic history 1984-2013 142 ;	
7. Bibliography and overview 187 .	
V The regulatory framework	
1. Developing and maintaining an effective regulatory state	190
2. Better regulation principles and institutional structures	192
VI Appendix and Index (p160)	
Searching by linear programming – the ‘Simplex’ method	197

I Political Economy

The course text is set out in six sections to provide a guide to political economy, and meet the needs of readers in various ways, whether they are students or teachers. The first section gives a brief introduction to political economy as a subject, and focuses on two major examples of work in the field of political economy. The second section sets out the foundations for economic analysis and accountancy, and includes two applied examples, one on controlling climate change and one on management control. The third section looks at six analytical examples to illustrate a range of technical problems which might need to be considered in practice. The fourth section looks at the UK experience of ‘independent economic regulation’ of privatised utilities and network industries since 1980. The fifth section provides an overview of the means by which political economy can be used to maintain institutional and managerial effectiveness through the regulatory framework. The last section is an appendix example on linear programming.

1. Introduction – The Scope of Political Economy

The term political economy speaks for itself, taken the Oxford English dictionary definitions of the words. Political means ‘relating to the government or public affairs of a country’, and economy means ‘the state of a country in terms of production and consumption of goods and services and the supply of money’. The dictionary definition of the two words together as ‘political economy’ simply refers to it, however, as ‘economics as a branch of knowledge or academic discipline’, and so we have to turn to the definition of economics to take us back to the literal translation. Economics is ‘the branch of knowledge concerned with the production, consumption and transfer of wealth’.

Political economy is therefore a generic subject which applies to the study of all countries, and to their relationships. Different countries will rule and manage themselves in different ways, and according to different beliefs, and countries will change over time. Some will be dictatorships, some communist and some market economies, but all will be equally open to analysis by political economy, as it is not a ‘value system’ but a method of inquiry. Some will believe in Communism – a social system characterised by the absence of classes and by common ownership of the means of production and subsistence – and some will believe in Capitalism – an economic and political system in which a country’s trade and industry are controlled by private owners for profit, rather than by the state.

In practice, most countries are somewhere in between, even though some may be trying to achieve a pure capitalist or communist outcome. Such a condition is reflected in the Marxist/Leninist doctrine of revolutionary struggle towards the goal of communism and socialism. At the other extreme, it would be rare to find a pure capitalistic state. The general tendency in the modern world is to be liberal, which means respectful and accepting of behaviour or opinions different from one’s own – and favourable to individual rights and freedoms. In all cases there is therefore a role for the state.

The role of political economy is therefore to support study, debate and the development of conclusions based on facts and coherent lines of argument. To represent its wide scope efficiently, it is sensible to link the idea of the ‘people’ with the ‘state’ by distinguishing private interest from public interest, and to recognise the roles of division of labour and division of responsibility in understanding how markets work and states are organised. To focus on these, and to reflect how things change over time, we will focus on two examples of political economy. The first will be Adam Smith’s *The Wealth of Nations*, published in 1776 (with its focus on self-

interest and the division of labour), and the second, published over 200 years later, on a Parliamentary inquiry into the ‘regulatory state’ (a new term of description) by the Constitution Committee of the House of Lords, given that the United Kingdom, along with most other countries, had moved from the pre-industrial state of Adam Smith to the modern state of today, and with the on-going debates about the division of responsibilities between people, families, communities and governments.

What the Committee had to say is relevant to any country, and draws attention to the great value of economic history in understanding how countries have, might or should develop. Great Britain has seen a long list from which to draw, whether it be the activities of village halls, the church, the philanthropists, the poor houses, public ownership or the welfare state. The modern ‘welfare state’ has emerged in part because of the transfer from predominantly rural to urban living in towns and cities, occasioned by the industrial revolution. This meant that the needs of the people could no longer be provided by the traditional roles of family, village and community life.

Before the two examples, it is interesting to note that the dictionary classifies the term political economy as ‘dated’. The term dated means that ‘it is no longer used by the majority of English speakers, but still encountered, particularly by older generations’. It is not that political economy has been found less useful, but that the specialist disciplines which lie beneath it have grown in strength and presentation. The list is a long one: from economics, law, accountancy, psychology, politics, social science and so on. An example is that even a university such as University College London (UCL) which had a Dept of Political Economy in 1970 with courses in pure and applied economics now has instead a Dept of Economics. These subjects are all needed, but one great advantage of political economy is that it is an ‘umbrella’ which should help to ensure that all the perspectives can come together to fully study the state of a nation, or particular policy problems. Too often specialist disciplines become inaccessible to the wider audience. The perspective was well put by ‘Schumpeter’ writing an article in *The Economist* (February 2011) on ‘Beyond Economics’ concerning the problems of political risk, which ended “It turns out that political economy is a much more complicated subject than its trendy modern offspring, economics”.

2. The Wealth of Nations – Adam Smith 1776

Why start with the *Wealth of Nations*? The 1976 edition, published by the University of Chicago Press, 200 years after Adam Smith’s ‘*An Inquiry into the Nature and Causes of the Wealth of Nations*’, refers to it “as the first scientific argument for the principles of political economy, it is the point of departure for all subsequent economic thought”.

The scope of the *Wealth of Nations* in its five books covering over 1000 pages makes it clear that it is a work on political economy:

- Book 1 – Of the Causes of Improvement in the productive Powers of Labour, and of the order according to which its Produce is naturally distributed among the different Ranks of the People
- Book 2 – Of the Nature, Accumulation, and Employment of Stock
- Book 3 – Of the different Progress of Opulence in different Nations
- Book 4 – Of Systems of Political Economy
- Book 5 – Of the Revenue of the Sovereign or Commonwealth.

The editor’s introduction to the 1976 edition suggests that it might have been called ‘*Principles of Political Economy*’ but notes that we should “remember that the term in 1776 was still a very

young one, and had recently been used by James Steuart in his book from the same publisher, ‘An inquiry into the principles of political economy: being an essay on the science of domestic policy in free nations’, 1767”.

The Wealth of Nations encompasses the ideas of both public and private interests, and how they might best be served, and reconciled, supported by analysis of human behaviour and economic history across nations. Nevertheless, for most he is probably best remembered for his passage which states that self-interest is in all our interests.

“But man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-interest in his favour, and show them that it is for their own advantage to do for him what he requires of them.....It is not from the benevolence of the butcher, the brewer, or the baker, that we can expect our dinner, but from their regard for their own interest”.

This reflects the division of labour and the place of the market which is at the heart of his analysis. The first sentence of book 1 states “The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgement with which it is anywhere directed or applied, seems to have been the effects of the division of labour”.

Adam Smith set out in the Wealth of Nations his theory of how markets work, how goods are sold to the highest bidder, and how the quantities of goods produced are governed by their costs and selling prices. His concept of stable and effective markets which increased the wealth of all nations was intended to be dynamic, and not static, drawing on the need to find the ‘natural’ price for any goods or services. It would be a mistake, however, to think that the Wealth of Nations is only about competitive markets and survival of the fittest, without regard to the rest. His purpose was also to demonstrate what ‘governments’ should best do in the public interest, showing that the policy of free trade was better than protectionism. Governments have a role because self-interest is related to public interest, so that government is necessary in any country to have regard for the provision of those public interests. In Adam Smith’s time, this would have meant the role of the ‘Monarchy’, and the tasks and resources of the Sovereign. Adam Smith sets out what forms of taxation might best supply the resources to provide for the public interest.

The first sentence of the introduction to book 4, Of Systems of Political Economy, underlines a clear view of the interrelation between public and private interests. He says “political economy, considered as a branch of the science of a statesman or legislator, proposes two distinct objects: first, to provide a plentiful revenue or subsistence for the people, or more properly to enable them to provide such a revenue or subsistence for themselves; and secondly, to supply the state or commonwealth with a revenue sufficient for the public services. It proposes to enrich both the people and the sovereign”. All of the concerns that are found in current political debates, whether to do with public service, or a fair society, can fairly be said to be found in the Wealth of Nations. Much is different from the time of Adam Smith – whether that be railways, air travel, digital technology, or the extent of the welfare and regulatory state – but the underlying role of political economy in helping to develop understanding and coherent policy in managing our lives effectively remains.

It should be noted that Adam Smith preceded the Wealth of Nations with his first book, ‘The Theory of Moral Sentiments’, which is concerned with the sentiment of sympathy. He says that

we are “interested in the fortune of others, and renders their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it”. The Wealth of Nations can be said to be a seminal alliance of politics, philosophy and economics (PPE). It should be no surprise to find, for example in Great Britain, that the state has always regarded a degree in PPE from Oxford University as a suitable training for those desiring to enter the higher levels of the civil service and public life.

3. The Regulatory State: Ensuring its Accountability – UK House of Lords 2004

Two hundred years after the publication of the Wealth of Nations the UK was just about to enter a new period in the design of the state and its governance. The last two centuries had seen the development of the four main essential service industries, water, energy, transport and communications, and these utilities and network industries became, after the end of the second world war in 1945, publicly owned, state enterprises. The state also provided the bedrock of education and health services. The public sector had become an equal partner alongside the private sector, and the institutions of the state were known and expected to meet the needs of the citizens for essential services, whether in education, health or infrastructure.

Alongside the public and private sectors, there is also the ‘third sector’, that is charities which run on a not-for-profit basis. The division of responsibilities between these three sectors is an important issue for political economy. The institutional structure of the state had also developed to include not just Departments of State, but other agencies, such as the Environment Agency, and a plethora of Quasi Autonomous Non Governmental Organisations (Quangos). The idea of independent (ie, non-political) regulation also became common outside the utility and network industry sectors, with bodies such as the Financial Services Authority (FSA) or the Office for Standards in Education (Ofsted).

Margaret Thatcher came to power in 1979 as the head of a new Conservative government, and one of its key objectives was to ‘roll back the frontiers of the state’ (but not necessarily its reach), which it decided had become too large and too inefficient. The providers needed to face the discipline of competition. The particular problem with the network, infrastructure industries however was that they provided essential services and often were ‘natural monopolies’. Privatisation of these industries, if it was to succeed and be sustainable, had to be accompanied by regulation, but at the same time the regulators had to be protected from day to day interference by politicians and government ministers. The state experiment begun in 1984 with the privatisation of British Telecommunications (BT) and the setting up of the first ‘independent economic regulator’, the Office of Telecommunications, known as OFTEL. This model was followed through with successive privatisations, although the structure of privatisations changed to promote more competition, notably in electricity and rail.

In these cases the industries were restructured prior to privatisation in order to separate the natural monopoly element in the supply chain, and so allow the other parts to be competitive. In the electricity industry it is the high voltage transmission and distribution networks, and in the railways, the track and signalling networks, in effect, the infrastructure. The competitive parts in electricity were generation and supply; in rail, the train services across the tracks. In rail, the passenger rail services were privatised with competition ‘for the market’ in particular regions of the country, and train operating companies (TOCs) who won the bid would be franchised to run the service for a specific period of time. The sequence of privatisation was:

1984 British telecommunications (BT) accompanied by OFTEL

1986 British Gas (BG) accompanied by OFGAS

1987 British Airports Authority (BAA) accompanied by the Civil Aviation Authority (CAA) and the Monopolies and Mergers Commission (MMC)
1989 Water and Sewerage Authorities accompanied by OFWAT
1990-1995 Electricity industry accompanied by OFFER
1992-1997 British Rail accompanied by ORR

When New Labour came to power in 1997 under Tony Blair, the emphasis changed to ensuring that the regulators, and not just the regulated, did a good job, and that consumer interests were properly represented with their own organisations. Over time the regulatory authorities have changed their names and functions in order to improve the governance of the regulatory state. Given convergence through digital technology, OFTEL was merged with others to form OFCOM, the Office of Communications. Gas and electricity were combined, given the introduction of competition to various parts of the supply chain, so that Offer and Ofgas became OFGEM, the Office of Gas and Electricity Markets. The independent economic regulators were based on giving statutory power to an individual, the 'Director General', but concern over power being too concentrated in the hands of one person, meant New Labour introduced the change to a 'board structure' with a Chairman and a Chief Executive. This is why ORR now stands for, not the Office of the Rail Regulator, but Office for Rail Regulation. The MMC was replaced by the Competition Commission.

The period of this state experiment 1979-1997, and thereafter, has been controversial because many believed in state owned industry, and particularly employees who were public service workers. The political economy of the privatisation and regulation of essential service network industries remains an important area of study for countries all across the world, and the 'regulatory state' has become an important term of reference for modern government. The key to understanding this framework is the term 'public services privately provided'. The experience of this form of state governance in the UK has meant that the initial tensions have been reduced, but they are appearing again now that privatisation of the Post Office is in prospect and the National Health Service (NHS) is perceived to be under attack from the new coalition government of Conservatives and Liberal Democrats formed in 2009 under David Cameron and Nick Clegg.

The Inquiry by the Select Committee on the Constitution in the House of Lords as its 6th report of the session 2003-04 has therefore been very timely. Entitled *The Regulatory State: Ensuring its Accountability*, and published in three volumes, given the large amount of oral and written evidence it received, it provides a good example of the use of political economy in practice. A selection from the text in chapter 1 'Introduction and Summary' of volume 1, Report, give a clear sense of its purpose. "Our starting point is that regulation is a means to an end, not an end in itself. Regulation can only be in the public interest where it serves a clear purpose" (4). "We conclude that regulators should be accountable for cost-effective regulation which meets rational, well-defined objectives" (8). To provide a proper discipline on regulators "The three key elements we identify are:

- * the duty to explain,
- * exposure to scrutiny,
- * the possibility of independent review (of their decisions)" (9).

From this comes the key statement that "effective regulation requires effective accountability" (10).

The reason for wanting effective accountability is because "accountability is a control mechanism which is an integral part of the regulatory framework" (10). An example of this is

that “The preparation of regulatory impact assessments (RIAs) is an important discipline on regulators” (10), and that “These RIAs need to be conducted retrospectively, as well as prospectively, to ensure that cost-effectiveness is constantly under review” (10). The role of Parliament is also examined: “It is not just a question of the answerability of regulators to Parliament, but also one of the duty of Parliament to ensure that its scrutiny is effective” (13). In this respect, their conclusion is clear “There is no means of establishing a coherent overview of the regulatory regime operating within the United Kingdom. We believe there should be.” (13).

Out of their 24 recommendations (most of which the government accepted in its response), a few have a particular resonance with political economy:

- * **First** “Independent consumer bodies should be obliged by statute to engage in open meetings and conduct regular surveys of consumers” (16,1).
- * **Second**, with regard to the board structure of regulators, “To ensure that there is no loss of accountability we recommend that boards designate one of their number as the public face of the regulator in order not to lose engagement with the public and to perform the role of confidence building and understanding” (16,3).
- * **Third**, “The Better Regulation Task Force (BRTF) should review its principles of good regulation to ensure that the principles of coherence, objectivity and rationality of approach are incorporated and signalled to the wider public” (16,10).
- * **Fourth**, “There must be a much stronger communication of the ‘whole of government’ view of regulation. We recommend that the government appoint a lead Department to be responsible for promoting effective regulation in practice, thereby co-ordinating the various roles currently played by a number of departments, including HM Treasury, DTI, The Cabinet Office and the Office of the Prime Minister” (16,11).
- * **Fifth**, “The move towards self-regulation should be encouraged and co-regulation should, where appropriate, be used as a preliminary to it” (16,13).
- * **Finally**, “Regulators should have a statutory duty to have regard to the principles of good regulation and effective accountability” (16,14).

Having provided a brief historical context with these two reports on political economy in practice, we can now move to the foundations of economic analysis and accountancy, following a list of the range of terminology which will be used throughout the book, for either principles or practice. It should always be remembered that whilst the main principles of economics and accountancy may remain fairly constant, the context of their use in political economy may often change. In particular, institutions are very prone to change, particularly when there is a change of government. In the UK, for example, the Office of Fair Trading (OFT) and the Monopolies and Mergers Commission (MMC), now the Competition Commission, are to be merged into the Competition and Markets Authority (CMA) by 2014; the Financial Services Authority (FSA) has been split into two parts, the Financial Prudence Authority, being the responsibility of the Bank of England, and the Financial Conduct Authority (FCA) supervising products and sales by banks; the Audit Commission for local bodies has been abolished in favour of private auditors for local authorities, and the National Audit Office (NAO) has new responsibilities for a maintaining a Code of Audit Practice; there are new ‘inspectorates’ for constabulary and for hospitals; and the Office of National Statistics is introducing a new inflation index (the RPIJ) to complement the standard RPI index.

Terminology of Political Economy

Positive and normative economics; macro and micro-economics

Welfare economics, economic geography, environmental (‘green’) economics

Demand and supply curves; the price elasticity of demand
Economic cost and accounting cost; opportunity cost and 'shadow' prices
Compounding, discounting and annuities; net present value (NPV)
Competition: 'perfect', oligopoly, duopoly, monopoly, 'natural' monopoly

Total revenue, average revenue and marginal revenue
Total cost, average cost and marginal cost
Fixed and variable cost; short and long run cost; 'absorbed' fixed costs
Constant, increasing and declining economies of scale
Normal profit, supernormal profit, and economic profit and loss

Balance sheets, profit and loss accounts, and cash flow statements
Assets and liabilities, income and expenses, receipts and payments
Double entry method for recording accounting transactions
The 'break-even' point in production and sales
Shareholders funds, net assets and capital employed
Historical cost and current cost accounting; revaluing fixed assets
The three costs of business: operating expenditure, capital consumption (depreciation) and the cost of capital finance
The cost of capital: interest on loans, dividends to shareholders

Investment rules based on NPV and internal rate of return (IRR)
Ranking independent and mutually exclusive projects, capital rationing
Ranking and the opportunity cost of deferral (OCD)
Profit maximisation, setting marginal cost equal to marginal revenue
Optimising the allocation of scarce resources, long run sustainability
Trade-offs and optimisation with multiple constraints

Financial ratios; return on capital employed (ROCE), gearing and leverage
Cost-benefit analysis; balanced score cards and zero-based budgeting
Variance analysis; budgets to outturn; the 'adding up' rule
Optimal output and pricing; primary products and by-products
Good economic profit (performance); bad economic profit (monopoly power)
Market failure – abuse of monopoly power, externalities, public goods and inequality
The 'free-rider' problem in public goods for private supply

The regulatory state: institutional framework for effective governance
Financial reporting and the principal/agent theory
Auditing and financial reporting standards
Cost-effectiveness and value for money
Risk and uncertainty; the polluter pays and precautionary principles

II The Foundations

Economic analysis

Economics is a specialist subject which is a branch of political economy. One key distinction is between ‘positive’ and ‘normative’ economics. Positive economics is the empirical analysis of how the economy actually works, while normative economics is concerned with ‘value’ judgements about how the economy should be. It is concerned with interpreting the effects from positive economics and judging whether policies need to be put in place to amend certain economic outcomes. In terms of application, it divides quite naturally into ‘macro’ and ‘micro’ economics.

Macro and micro economics

The purpose of this is to divide small scale, detailed analysis from the analysis of economic systems as a whole. Microeconomics is concerned with examining one aspect of economic behaviour, for example, the markets for particular types of goods, or focusing on individual firms. Such analysis can be aggregated to provide a more complete picture, and the summit of such analysis is ‘general equilibrium’ theory. However, such complex analysis might make it impossible to focus on the key elements of the bigger picture, and macroeconomics overcomes this by focusing on interactions in the economy as a whole. Macroeconomics is concerned with ‘consumer goods’ as a whole, rather than specific types of consumer goods. Macroeconomics tends to focus on key national statistics, such as Gross Domestic Product (GDP) – the total output of goods and services in the economy; the aggregate price level – the average level of prices of the goods and services (which is a measure of inflation over time); and the unemployment rate.

For government policy, macro-economics is concerned with the management of the economy as a whole, incorporating such things as managing the money supply and controlling inflation, taxation policy, and managing the economy through public spending to avoid recession (notably known as Keynesian economics). Fiscal and monetary policy is a key policy component for all governments, and includes managing the exchange rate for the purposes of international trade and maintaining the balance of trade in exports and imports. Micro-economics is concerned with the level of the firm, and how they do, or should, behave in the context of competitive markets. The study of corporate finance sits well within micro-economics, but is very relevant to macro-economic policy.

Distinguishing economic and accounting costs – Opportunity cost

Our concern is not with the specialist subjects of macro and microeconomics, but with the foundations which underpin much of economic analysis; foundations which are necessary for solving problems and deciding on the best policy stance. These include the financial aspects of economics, such as discounting and investment rules, as well as the idea of ‘optimal’ outcomes, and the conceptual frameworks within which those are set. Given the role of money in exchange which underpins the operation of markets, and that time is an important factor in all contracts and policy decisions, we should start with present and future values, preceded by a section on ‘economic cost’.

Economics and accountancy are never far apart, but it is important from the start to focus on the term ‘opportunity cost’. The reason for this is that economics is concerned with the decision-making process, whilst accounting is concerned with recording the financial aspects of that decision-making. In this respect the definition of cost may differ. Resources, whether labour, materials or machines, can usually be used to produce different things, and accounting is concerned with recording the financial cost of the chosen use for the resources. However, economic decisions are made to choose what use to make of the ‘scarce’ resources. Defining

the economic cost of using a resource on a particular product therefore requires an analysis of what would be ‘foregone’ by not using the scarce resource for another product.

The definition of opportunity cost is therefore the value which the resources would have had if used to produce the next best alternative use. An example of the distinction between economic and accounting cost would be where 10 hours of skilled labour at £15 per hour is used to produce a product (A) which makes a profit of £100 for each sale. The accounted cost of the skilled labour is £15ph, but suppose the skilled labour could be used on another product which generated a profit of £150. The opportunity cost of skilled labour for product A should therefore include the additional profit foregone, which is $150 - 100 = 50$. Given product A uses 10 skilled labour hours, the opportunity cost per skilled labour hour would be $£15 + (50 \div 10) = £20$.

1. Present and future values

Financial analysis means dealing with a series of financial numbers, or cash flows, some of which are receipts, some of which are payments. Typical text book presentations are often concerned with an investment (a payment), from which there will be a run of receipts over time. To distinguish payments from receipts in a single line of cash flow, the payments might typically be shown in a bracket (). The time period between each is typically a year, but could be months, weeks or days as needed. Year 0 is now, year 1 is the end of year 1, and so on.

Year	0	1	2	3
(payments) receipts	(x)	x	x	x

It is important in economic and financial analysis to be able to reduce a cash flow series to a single number, and this requires a ‘discount rate’ which reflects the time value of money (a £1 available today is worth more than a £1 available at some time in the future). The discount rate (r) will depend on the analysis in question, given the analysis might relate to an individual, a company or the government. An individual will be concerned with their own ‘time preference’ rate; a company with the cost of funding its investments from shareholders funds or bank loans (referred to as the ‘cost of capital’; and governments will have their own ‘test discount rate’ (perhaps referred to as the social time preference rate or the social opportunity cost rate). The appropriate discount rate is a specialist field of study, but for this book we are only concerned with a given ‘benchmark’ test discount rate (r^*), and the results of calculations for different discount rates r .

Compounding

Present values and future values can be linked by compounding and discounting. If we have a sum of £ x ‘now’, then we can invest that to earn a return by the end of the year. This means we can derive the future value (FV) of a sum £ x now at the end of the year from $x(1+r)$. If that future sum was reinvested for a further year, then we earn interest on the original investment (the principal) plus the interest earned in the first year, that is interest on the interest, known as ‘compound’ interest. At the end of the second year the future value would be $x(1+r)(1+r)$. This can be continued for any number of years, t , and the future value would be $x(1+r)^t$, that is, x times $(1+r)$ raised to the power of t . Hence FV of $x = x(1+r)^t = y$.

Discounting

Given this, we can take a sum £ y in the future at time t as its future value and discount it to find its present value (PV). The present value can be found from rearranging the future value formula above to give the PV of $y = y$ times $1/(1+r)^t$. The term $1/(1+r)^t$ is the discount factor to be applied to the future sum in question. For example, given an investment of 100 now, its future value in two years time at a cost of capital of 10% is $100(1.1)^2 = 121$. The present value of 121 receivable in two years time is $121(1.1)^{-2}$, that is, 100. The discount factor (df) represented by $(1.1)^{-2}$ is $1/1.21 = 0.8264$.

Net present value

The present value of a series of receipts and payments is known as the net present value (NPV). The NPV can usefully be divided into the present value of its receipts (PVR) and payments (PVP) because $PVR - PVP = NPV$.

Annuity - calculation of the constant annual sum

A very useful result arises from cash flow series that are a fixed payment or receipt (a) each year because if $PV = a(df_1) + a(df_2) \dots a(df_n)$ then $PV = a(\sum df_s)$, therefore the fixed annual sum a is the present value of the cash flow series divided by the sum of the discount factors, that

is, $\frac{PV}{\sum df_s}$, given the number of years of the series and the interest rate. Note that given a cash

flow series with different cash sums for each year, annuitisation can be a very useful way to convert the present value of that cash flow series into an equivalent series of constant annual sums. Given the present value of the cash flow series, simply divide it by the sum of the discount factors over the life of the cash flow series. Equally, the value of an investment made today (C), can be converted into an equivalent annual sum for a given number of years t . This annual sum will incorporate two parts, principal and interest, the constant annual sum over the life of the investment which would repay both the investment (principal) and the financing costs of the investment (the cost of capital represented by dividends and/or interest).

Discount and annuity tables

Discount factors are a function of time (t) and the rate of interest (r), and given r and t , can be worked out or easily looked up in 'discount tables' which show the discount factors typically ranging from $r = 1 - 30\%$ and $t = 1-50$ years. Note in particular that $(1+r)^0 = 1$ (ie, when $t =$ now, then the cash flow at that time is undiscounted), and that when $r = 0$, then the discount factor is also 1. This is a useful result because the sum of an 'undiscounted' cash flow series is the same as the present value sum of a 'discounted' cash flow series at an interest rate of zero (because $(1+0)^t = 1$). The sums of discount factors can be looked up in 'annuity tables' for, as with discount tables, a set of years and interest rates.

The following table shows a selection of discount rates for different interest rates and years. These are often published to either three or four decimal places, and because these are 'rounded' approximations of the numerically correct discount rate, their use in practice may result in often minor 'rounding' errors. Take for example the discount rate for one year at 10%. To three decimal places it is .909 and to four decimal places it is .9091 whilst $1/1.1$ actually equals .9090909.....

Years/interest rates	Discount factors							
	1%	5%	10%	15%	20%	30%	40%	50%
1	.9901	.9524	.9091	.8696	.8333	.7692	.7143	.6667
5	.9515	.7835	.6209	.4972	.4019	.2693	.1859	.1317
10	.9053	.6139	.3855	.2472	.1615	.0725	.0346	.0173
20	.8195	.3769	.1486	.0611	.0261	.0053	.0012	.0003
30	.7419	.2314	.0573	.0151	.0042	.0002		

It can be seen that as the interest rates or years to be discounted rise, the discount rate to calculate the present value quickly becomes a very small number. The annuity tables are set out in a similar way, showing the sum of the discount factors for the number of years in question at a particular interest rate, which is the present value of £1 receivable for that number of years given the interest rate. For example, the annuity for 5 years at 10% would be 3.7907, the sum of .9091+.8264+.7513+.6830+.6209

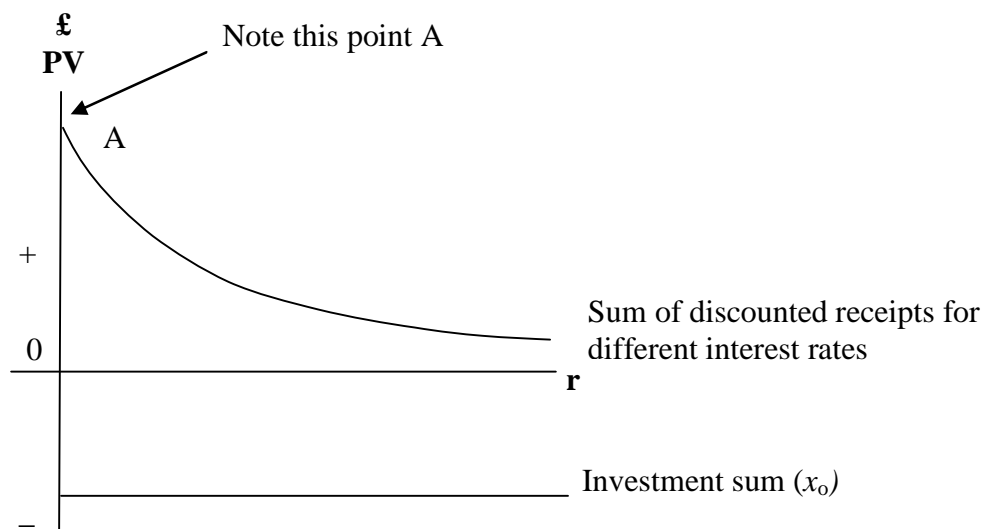
Present value charts and the Internal Rate of Return (IRR)

Focusing first on positive discount rates, the relationship between present value and the discount rate can be plotted as a graph. Note that point A is the present value of the series of receipts at a

zero discount rate. This is the same thing as the undiscounted sum of the series of receipts (ie, simply added-up). This will be a useful result for making a rough estimate of the internal rate of return (IRR). The curve shown is the discounted sum of the cash flow of receipts in question as the interest rate rises from 0 towards infinity. The sum of the discounted receipts will tend towards zero as the interest rate rises. This is because the discount factor $1/(1+r)^n$ will become a very small number as r increases. The investment sum x_0 which occurs 'now' in time remains a straight line because its discount factor is equal to 1 (ie, $(1+r)^0 = 1$).

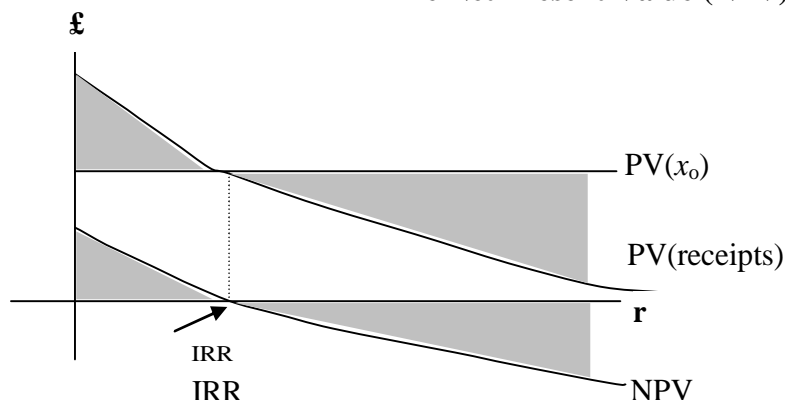
Normally the receipts and payments are plotted on the same axis so that any intersection point can be identified, which is the point where the present value of receipts (PVR) = the present value of payments (PVP). This is where $NPV = 0$. The difference between the present value of receipts and payments lines can be plotted as the net present value (NPV) line, as shown (ie, $PVR - PVP = NPV$). An equally important result is that the $PVR = PVP + NPV$.

Present value curve



The intersection of the NPV curve with the zero £s horizontal axis is known as the Internal Rate of Return (IRR), as shown in the diagram, and is the point where $NPV = 0$, and $PVR = PVP$. This is an important point in investment appraisal, as is knowing whether NPV is greater or less than zero. The definition of the IRR is therefore that rate of interest which makes the discounted cash flow series = 0.

The Net Present Value (NPV) curve



Constant prices – nominal and real interest rates

Inflation changes the value of money over time and has to be adjusted for in financial analysis if we are to compare things over time. We may earn twenty times as much as we did ten years ago, but prices may be twenty times higher. There are various indexes of inflation, such as the Retail Prices Index (RPI) and the Consumer Prices Index (CPI), which differ because they have different ‘baskets’ of goods on which the indexes are calculated. For example, some exclude housing costs. The choice of index to be used in analysis will depend on the purpose of the exercise (and is a specialist subject in its own right), but in this book we will simply refer to the RPI. Nominal discount rates include an allowance for the effect of inflation, therefore real discount rates exclude it. Hence, for example, if the real interest rate (r) is 5% and the inflation rate (i) is 10%, then the nominal interest rate is: $(1.05)(1.1) = 1.155$ or 15.5%.

Cash flow series can be prepared in real or nominal terms (ie, incorporating inflation into the forecasts of receipts and payments) and they must be discounted with the appropriate discount rate. In practice, if we are converting a nominal cash flow series (ie, incorporates inflation) into the price level of a particular year (eg, to create a constant price series at, say, year 2000 prices) then we use the retail price index (RPI) ratio for each of the two years in question to correct to a constant (year 2000) price series (the ratio may be a multiplier or a deflator as appropriate). So, for example, the nominal CF sum at 1990 prices in 1990 times the RPI index 2000 divided by the RPI index 1990 equals the 1990 sum in year 2000 prices. The total inflation rate over the period would be $(1+i)^t$ which equals the RPI multiplier for the period, therefore we can see that $i = \sqrt[t]{(\text{RPI multiplier})} - 1$

Interest rates for periods less than a year

Banks offer accounts with monthly and daily interest rates. These periodic interest rates have to be related to the underlying annual interest rates. If we say that our annual interest rate is r_1 , then a half yearly rate would be r_2 , and a quarterly interest rate would be r_4 , and so on. If an equivalent half-yearly compound interest rate (r_2) was required for an annual rate (r_1) then we would have: $(1+r_2)(1+r_2) = (1+r_1)$ or $(1+r_2)^2 = (1+r_1)$ therefore $r_2 = \sqrt{(1+r_1)} - 1$ and equivalently for quarterly rates, ie, $r_4 = \sqrt[4]{(1+r_1)} - 1$ and so on (ie, $r_n = \sqrt[n]{(1+r_1)} - 1$).

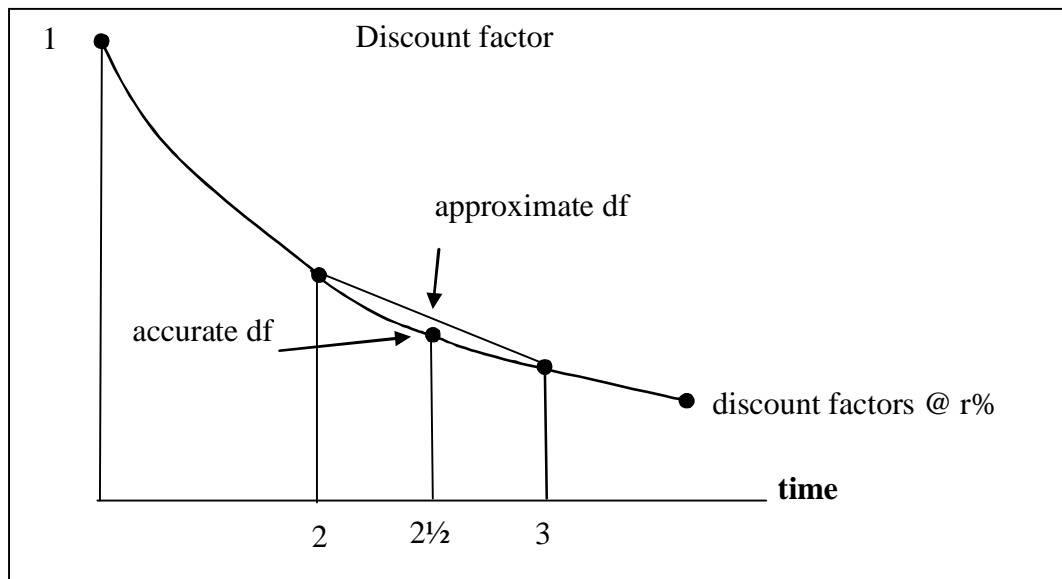
Cash flows receivable half way through the year (or more frequently)

If account needs to be taken of this, then the discount rate needs to apply to receipts or payments occurring at sometime during the year. Two methods are available:

- *approximate*: calculate the discount factor as the average of the opening and closing discount factors for the year (ie, the normal annual discount factors)
- *accurate*: so, for example, if we are interested in the discount factor for a cash flow in three and half years time we would have a denominator of $(1+r)^{3\frac{1}{2}}$ which is the same as $(1+r)^3 \times \sqrt{(1+r)}$ and if you wished to calculate quarterly, eg, $2\frac{3}{4}$ years time, then it would be $(1+r)^{2\frac{3}{4}}$ which is $(1+r)^2 \times \sqrt[4]{(1+r)}^3$

The chart shows the relationship between the approximate and accurate discount factors. In practice these part year discount factors are found using the ‘function’ routines in computer spreadsheets or in scientific calculators.

Accurate and Approximate part year discount factors



2. The Standard Investment Rule

The text book standard investment rule requires a 'test' discount rate r^* which is used to make comparisons for decision making. For company analysis, r^* can be taken as the 'cost of capital' to the firm. The chart shows the position for two different costs of capital, r^*_1 and r^*_2 .

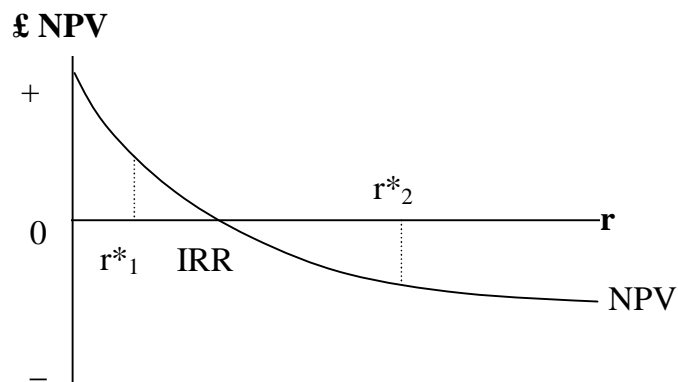
The NPV chart shows clearly the standard investment rules (based on the idea that NPV is a measure of the wealth increase or decrease from a decision).

Invest if: $NPV > 0$ or $r^* < IRR$

Don't invest if: $NPV < 0$ or $r^* > IRR$

The NPV curve shown is for a 'normal' cash flow series (an outgoing followed by a series of receipts, ie, $(x) \times x \times x \times \dots$). More complicated cash flow series can cause problems for IRR analysis, as there can be more than one value of the discount rate r which makes the $NPV = 0$. We return to this problem and its reconciliation in the analytical examples.

The NPV curve and the IRR



Solving for the IRR

The IRR for a cash flow series cannot generally be solved directly, but requires an iterative sequence. Computers and calculators are programmed to do this for you, once you enter the cash flow series. It is useful to know, however, that an iterative procedure can be done by hand quite simply, and to a reasonably accurate estimate. This is because you normally already have two points on the NPV curve for a cash flow series. First, the NPV at a zero interest rate (the arithmetic sum of the cash flow series) and, secondly, the NPV discounted at the cost of capital r^* (as this would normally be required for any financial and economic analysis).

Estimating the IRR iteratively where the cash flows are not constant annual sums

It is solved by an iterative routine which converges on the solution. However, we can see from the NPV chart below the sequence of steps which enable quick estimation:

Step 1: calculate the NPV @ the cost of capital r^* (point B)

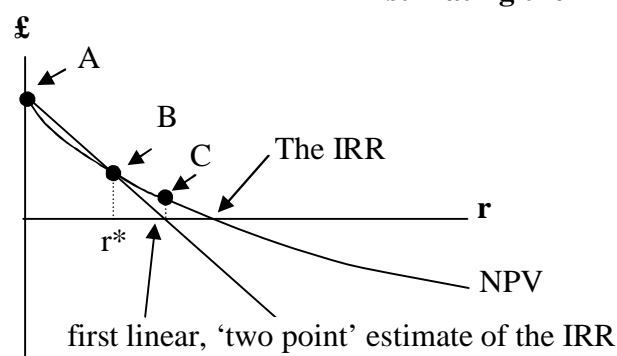
Step 2: calculate the NPV at a zero interest rate by summing the cash flow series (point A). This point shows that the sum of the cash flow series is a positive number (if it was negative the IRR would be negative)

Step 3: estimate the IRR by straight line extension

Step 4: calculate the NPV based on the first estimate of IRR (which would be point C)

Repeat by linear extension from points B and C and at this point the estimate of IRR is probably accurate to within a half % point.

Estimating the IRR by linear extension



Estimating the IRR from an annuity cash flow

Where the cash flow of receipts is a constant annual sum (a), and given the initial investment is C , then because the IRR is the rate of interest r which sets the $NPV = 0$, then $C = a(\text{sum of the discount factors})$, and the sum of the discount factors $= C/a$. You can solve for the IRR by looking for the rate of interest r in the annuity tables where the (Σdf_s) is equal to C/a . An example: $C = 100$; $a = 20$ receivable for 7 years. Hence $C/a = 5$. The IRR is therefore approximately 9% because from the annuity tables, £1 receivable each year for 7 years has a present value of £5.033 at an interest rate of 9% per annum.

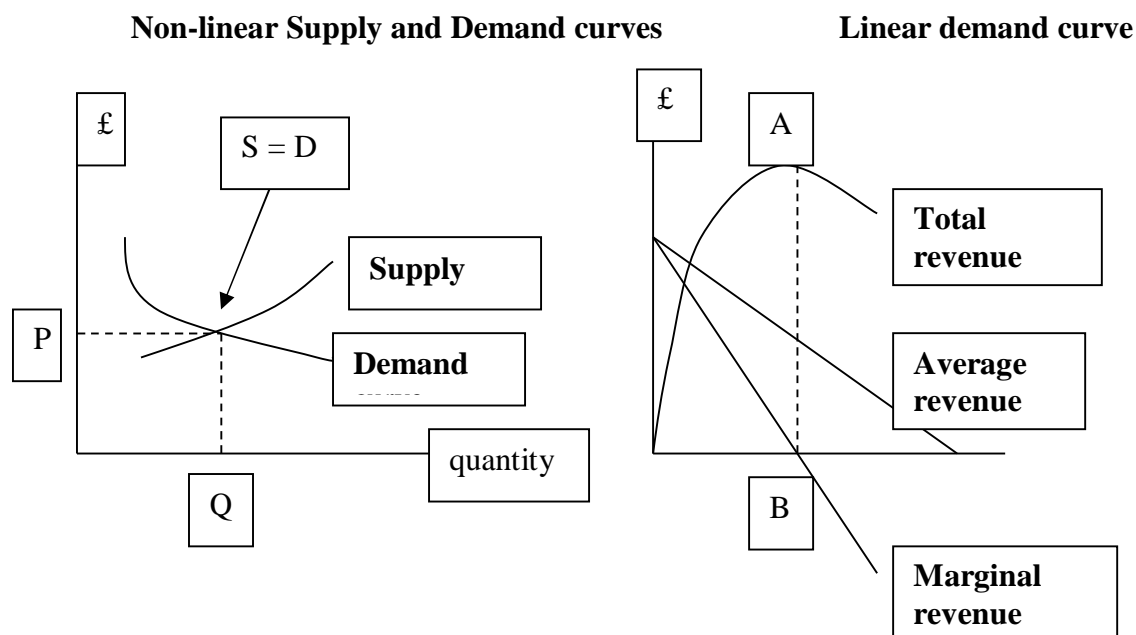
3. Demand and supply curves

Demand curves link price (p) and quantity (q) and are a dominant feature of text books and economic analysis. The 'downward' sloping demand curve (whether for an individual organisation or the industry as a whole) reigns supreme, given the assumption that the higher the price, the less will be bought, and the lower the price the more will be bought. Demand and supply curves bring together the actions of consumers and producers to a point where demand equals supply. The upward sloping supply curve reflects the fact that if the price is higher, more will be supplied. There is nothing that makes a demand or a supply curve fixed, and their estimation is an empirical subject. The supply curve will depend not only on the price of the

commodity, but the prices of all other commodities, the prices of its factors of production, new technology, and even behavioural aspects, such as the goals of its producers. Supply curves, like demand curves, can shift or change shape. The affect of substitutes for, and complementary products for, a particular commodity will be important influences. The typical representation of a demand and supply curve is shown below.

The 'demand' curve reflects quantity as a function of price, $q = f(p)$, but the convention is to chart demand curves as though price is a function of quantity, $p = f(q)$. Nevertheless, it is not a problem, but helpful. More important is that demand curves are typically shown as non-linear curves, as in the first chart below. Analysis however is simplified when linear demand curves are used, and the relationships between average, marginal and total revenue curves can be simply shown, and the optimal, profit maximising point identified for the production decision rule, which is to set marginal cost (MC) equal to marginal revenue (MR), ie, $MR = MC$.

If the linear demand curve is $p = a - bQ$, then the total revenue curve ($P \times Q$) is a quadratic, $aQ - bQ^2$. The marginal revenue is the small change to total revenue (TR) for a small change in the quantity Q . By differentiation (dTR/dQ) we get from the total revenue quadratic $a - 2bQ$. Clearly the marginal revenue curve falls twice as fast as the demand curve, which is a useful result. But it returns us to what the demand curve really is financially. If total revenue is P times Q , then average revenue is PQ divided by Q , which is P . The demand curve as plotted above is therefore also the average revenue curve. The important results can clearly be shown on a chart. For a linear average revenue (demand) curve, the marginal revenue curve falls twice as fast, and the marginal revenue curve is zero at the point which is the maximum of the total revenue curve. The marginal revenue curve becomes negative as total revenue falls. As the second chart below shows the three curves combined in order to illustrate the various 'relationship' points (A and B), it would require in practice two vertical (£) axes to accommodate the different sizes of financial numbers relevant to total revenue curves compared to average and marginal revenue curves (eg, AR is £5 per unit and TR is £5000 for sales of 1000 units).

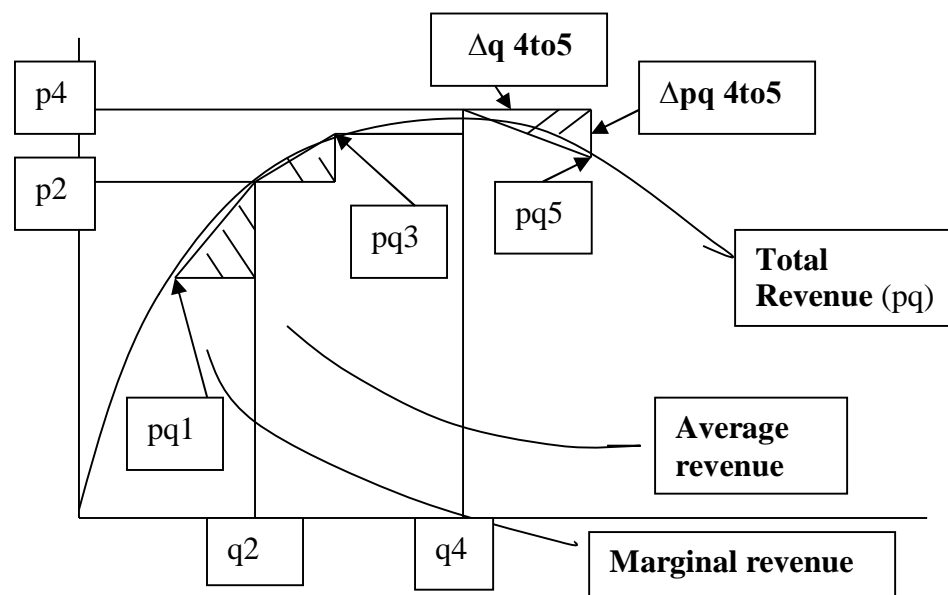


Non-linear marginal revenue curves

We have set out the simple representation of the total, average and marginal revenue curves which result from a linear average revenue curve. If you have a non-linear demand curve as is typically shown in text books, then you have a marginal revenue curve which is non-linear. If the demand curve does not reflect a simple algebraic equation then an estimate of the marginal revenue curve could be made by incremental analysis. If, for example, a company's marketing team provided you with a series of estimates of the quantities which are expected to be sold for a number of different prices, then the incremental change in quantity (ΔQ) for the incremental change in total revenue (ΔTR) can be found from the differences between them. The total revenue curve points are $P.Q = TR$ and the incremental revenue (ΔTR) = $TR_{n+1} - TR_n$. The associated incremental revenue (ΔQ) = $Q_{n+1} - Q_n$. The 'estimated' marginal revenue between each of the series can then be calculated for each from $\Delta TR/\Delta Q$. The diagram illustrates this.

The slope of the line (hypotenuse) for the incremental triangles shown is a reasonable estimate of the marginal revenue of the total revenue curve at the mid-point between Q_1 and Q_2 and so on. The average revenue points are plotted consistent with the price and quantity combinations given in the total revenue curve, whilst the marginal revenue curve estimates are plotted at the mid-points in the series of the quantities given. A reasonable line can be drawn between the points to create a continuous curve. Such calculations can be a useful exercise to bring home the understanding of the relationship between total, average and marginal revenue curves.

Estimating non-linear marginal revenue curves



The above can be illustrated with a simple worked example, assuming that your marketing department has provided you with five forecasts of some price and quantity combinations, and that the unit cost of producing the good in question is £3.

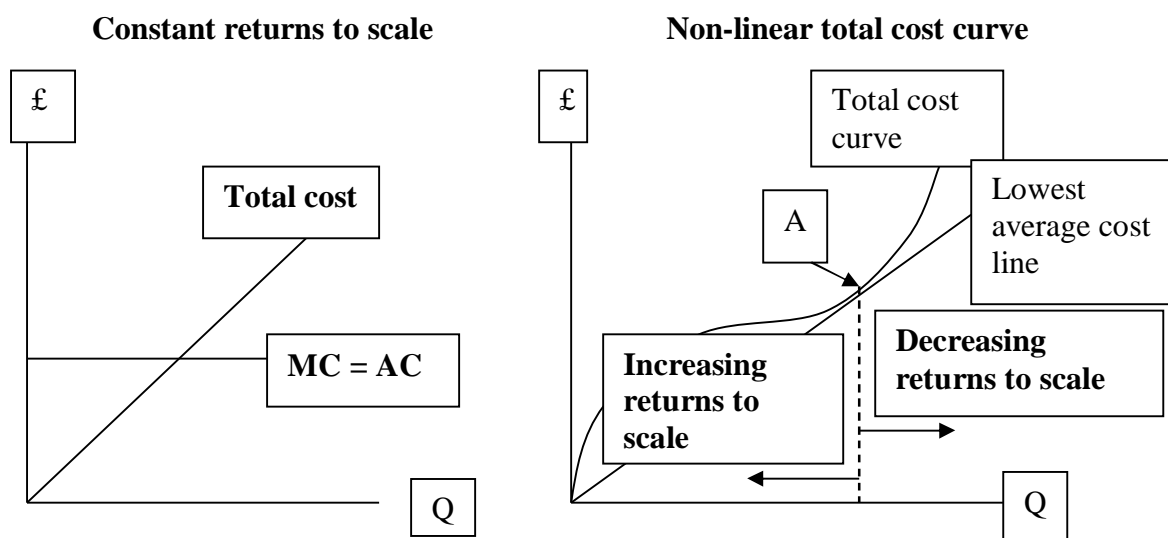
P	Q	TR	ΔTR	ΔQ	$\Delta TR/\Delta Q = \text{average MR}$
20	100	2000			
17	120	2040	+40	20	$40/20 = 2$
12	200	2400	+360	80	$360/80 = 4.5$
7	350	2450	+50	150	$50/150 = 0.333$
5	450	2250	- 200	100	$-200/100 = -2$

As expected, the price falls and demand goes up. The total revenue is highest at a price of £7, but note that the marginal revenue at £0.333 is less than the cost of supplying the goods at £3. Given profit maximum is where $MR = MC$, then it is likely, having drawn the curves on a graph, that the best price will be around £10 with sales around 250.

Cost curves

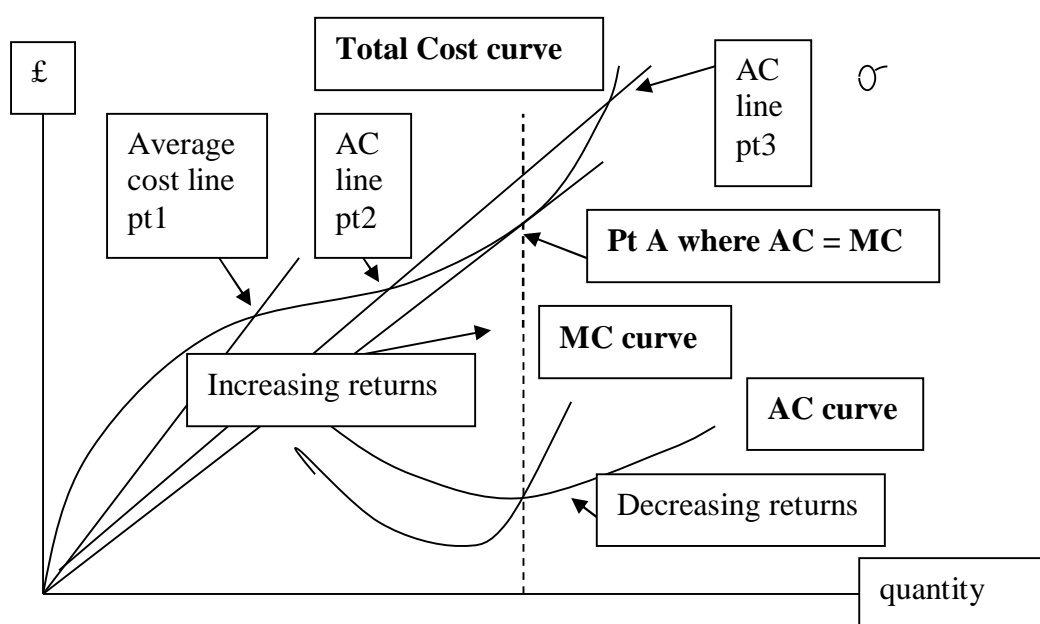
Cost curves, as with revenue curves, have total, average and marginal curves, and it is useful at this point to introduce the idea of ‘returns to scale’. Returns to scale are the relationship between an extra quantity produced and the cost of producing that extra quantity. Where there is an equal proportionate increase in the quantity and the cost, then we have ‘constant returns to scale’, which is helpfully linear. The chart shows the total cost curve as a straight line from the origin, and the slope of that line will determine the average and marginal cost per unit of quantity, which are equal, shown as a horizontal line on the chart.

Where the total cost curve is not linear, then it is subject to either ‘increasing’ or ‘decreasing’ returns to scale. Increasing returns may be because producing a large quantity of something means advantage can be taken of economies of scale, and the average cost falls as more is produced. This can be important in infrastructure investment decisions, for example, when deciding what size reservoir to build, as a dam for two million cubic metres might cost less than twice the cost of a dam for one million cubic metres. Decreasing returns to scale might arise because only lower quality resources might be available to produce a higher quantity, or resources are scarce and competition for additional amounts raises prices. The diagram of a total cost curve illustrates these. Point A is the point on the total cost curve at which the average cost is at its lowest point.



The important thing to notice from such a non-linear total cost curve is the affect on the average and marginal cost curves. The slope of the total cost curve is clearly changing as the quantity supplied increases, which gives us the marginal cost curve. The changing slopes of the lines drawn from the axis to different points (1,2,3) on the total cost curve reflect the changes in the average cost curve. Note that the average cost curve equals the marginal cost curve at point A on the total cost curve, the cross over point for the marginal and average cost curves and the point where increasing returns to scale become decreasing returns.

Non-linear total cost curves and their AC and MC curves



Profit maximising production

Profit (or surplus) is the difference between the total revenue and the total cost, and profit maximising is to find the production (sales) point which maximises the difference. The decision rule is to set marginal cost equal to marginal revenue. To increase production beyond that point would mean MR was less than MC, therefore not worthwhile. We need to introduce the marginal cost and total cost curves into our chart. For simplicity we will assume a constant marginal cost per unit of quantity, but include within it the ‘three costs of business’, which are:

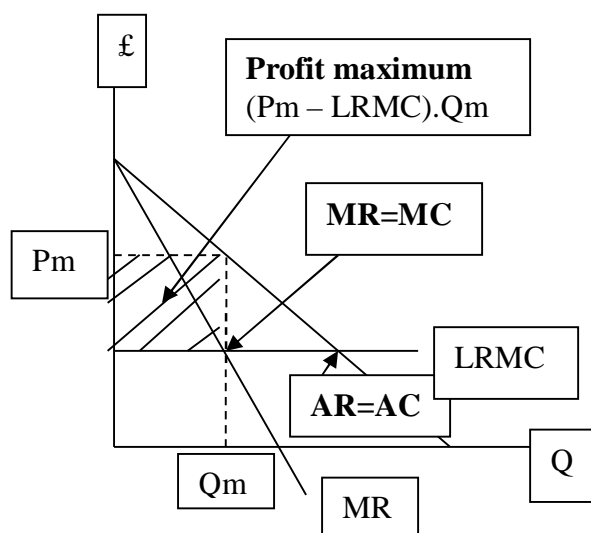
1. Operating cost (non-capital expenses such as wages, electricity etc)
2. Cost of capital consumption (depreciation of capital assets)
3. Cost of capital finance (cost of funding the purchase of capital assets)

The two costs of capital assets can be made into a constant annual sum by annuitisation, and therefore this can be said to be a measure of long run marginal cost (LRMC). The intersection of the marginal cost and the marginal revenue curves gives us the optimal production point, and the price necessary to result in that quantity being demanded is found from the average revenue (demand) curve, as shown in the chart. The profit is maximised and in finance terms is equal in this case to $(P_m - LRMC)Q_m$. In terms of total revenue and total cost the chart would show the following. The distance between points A and B is the maximum difference between the total cost and revenue curves as Q increases.

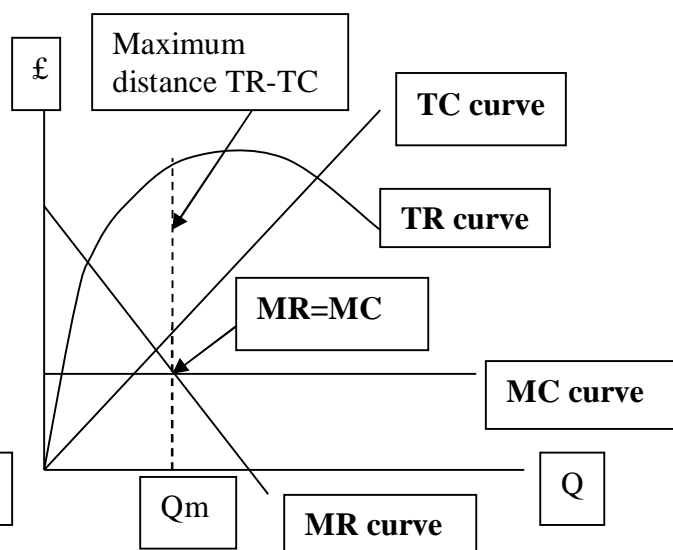
4. Competition, choice and monopoly profit

We have shown the standard downward sloping demand curve to illustrate the relationships between marginal, average and total revenue curves, and the profit maximising point. With competition and choice, however, you would expect to see a different demand curve for a competitive supplier than a monopoly supplier. This is because competition makes it difficult to have prices much higher than cost (otherwise consumers would buy from other suppliers). In the charts below the profit maximising price is quite substantially above cost, so it is a monopoly supplier. For our demand curves we therefore introduce the idea of ‘elasticity’.

Profit maximising AC and MC curves



Profit maximising TR and TC curves



Elasticity of demand

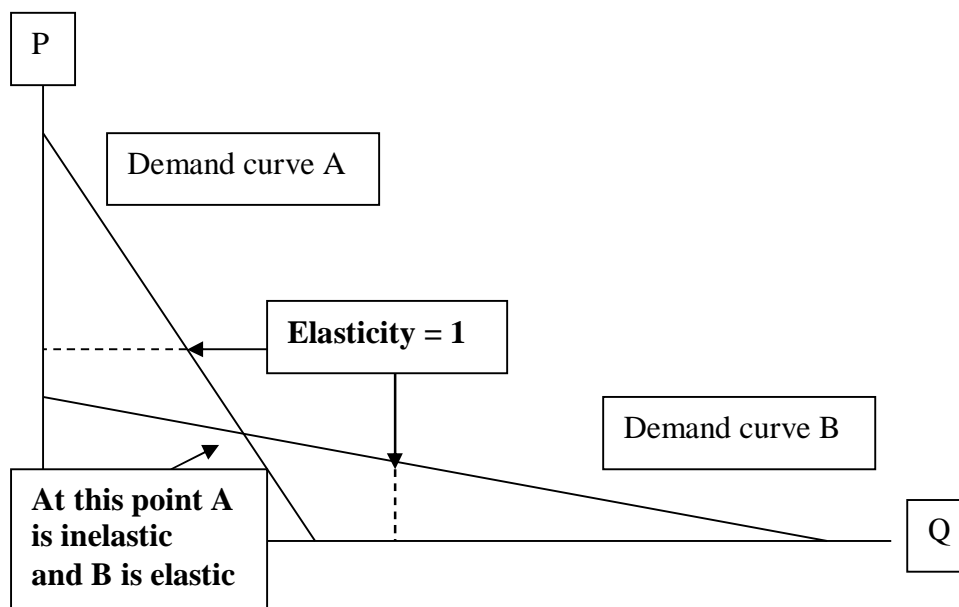
Elasticity reflects the % change in quantity for a % change in price. Elasticity of demand is an important concept in economics, and to business management, because it is a measure of the 'sensitivity' of the quantity demanded to the price. With competition and choice, you would expect more 'elastic' demand curves, that is, if a supplier tries to charge higher prices than their competitors, you would expect a rapid fall in sales. Their demand curve is more horizontal. For a monopoly supplier, where consumers have no choice, they are likely to have to pay the higher price within reasonable bounds, and the demand curve will be more 'inelastic', and tend more towards the vertical.

This is known as 'price elasticity of demand', but there are two others, the 'cross' price elasticity of demand, which measures the sensitivity of the quantity demanded of one good to changes in the price of a related good, and the 'income' elasticity of demand, which measures the sensitivity of the quantity demanded to changes in consumer income.

The price elasticity of demand is measured as the percentage change in the quantity of a good demanded divided by the corresponding percentage change in its price. This can be represented as $\Delta Q/Q \div \Delta P/P$, and by rearrangement is $\Delta Q/\Delta P \times P/Q$. For very small changes in price, the formula would calculate the 'point' elasticity of demand, that is, $dQ/dP \times P/Q$. In formula terms elasticity is a negative number because demand curves are normally downward sloping, and therefore a reduction in price is matched by a positive increase in quantity and vice versa. In presentation terms the negative sign is usually omitted because what is important is whether the elasticity is greater or less than 1. Demand curves as such are not elastic or inelastic because all demand curves have a range of elasticities depending on where the point is on the demand curve. The elasticity will fall from a large number where the price is high and the demand very low to a small number where the price is nearly zero and the demand very high. For linear demand curves, the elasticity will equal 1 at the mid-point of the demand curve (ie, an equal proportionate increase in quantity for a proportionate reduction in price).

The elasticity of demand has an important link therefore with the total revenue curve. When the elasticity is equal to 1, the marginal revenue is zero, and the total revenue is at a maximum.

The demand curve is therefore 'elastic' (greater than 1) when the total revenue is rising and 'inelastic' (less than 1) when the total revenue is falling. In text books two demand curves are often compared, and the one with the steeper gradient is referred to as inelastic, whilst the other one, with a more horizontal gradient, is referred to as elastic. This arises because at the cross over point for the demand curves, one is still in its 'elastic' half and the other in its 'inelastic' half. The chart illustrates this:



A worked example is useful to show the calculations for two such demand curves, being $P_A = 20 - 2Q$ and $P_B = 10 - 0.5Q$, and therefore $Q_A = (P - 20)/-2$ and $Q_B = (P - 10)/-0.5$. For demand curve A, given three price/quantity combinations, we would have elasticities (E) of :

P	Q	ΔP	ΔQ	P	Q	$\Delta Q/Q$	$\Delta P/P$	E
17	1.5	+1	- 0.5	18	1	0.5/1.5	1/17	5.66667
10	5	+1	- 0.5	11	4.5	0.5/5	1/10	1
4	8	+1	- 0.5	5	7.5	0.5/8	1/4	0.25

and for demand curve B

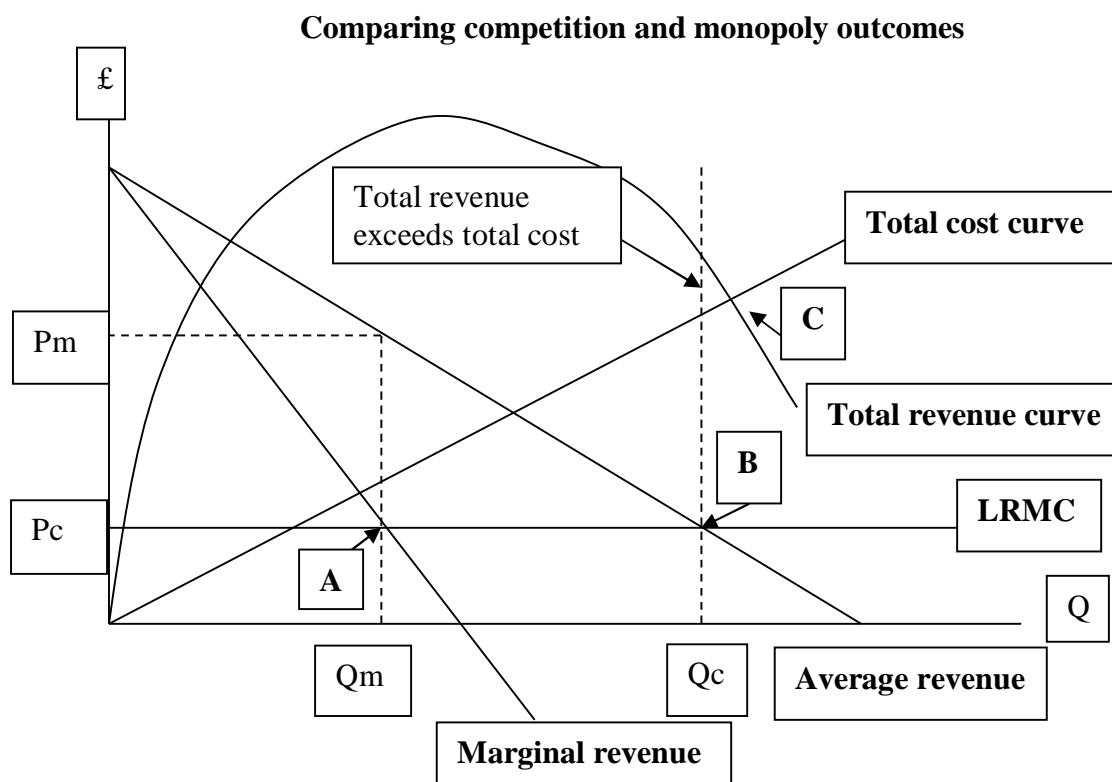
8	4	+1	- 2	9	2	2/4	1/8	4
5	10	+1	- 2	6	8	2/10	1/5	1
2	16	+1	- 2	3	14	2/16	1/2	0.25

Having shown that the elasticity of demand varies along the linear demand curve, it is useful to note that the incremental method (Δ) of calculating the elasticity at a particular price/quantity combination will provide the same result as from the point elasticity by differentiation approach for a linear demand curve. Take demand curve A above for example. To calculate $dQ/dP \times P/Q$, given $P = a - bQ$, then because $Q = (P - a).1/b$ the dQ/dP must be a constant. If the price/quantity combination was $P=12$ and $Q=4$, and $a = 20$ and $b = -2$, then $dQ/dP = 0.5$ and the point elasticity of demand would be $0.5 \times 12/4 = 1.5$ in the 'elastic' half of the demand curve. If the incremental method had been used then the calculations would have shown for three incremental changes in price ΔP (+2, -1 and -4) the following: $1/2 \times 12/4 = 1.5$; $0.5/1 \times 12/4 = 1.5$ and $2/4 \times 12/4 = 1.5$. The result is always two constants multiplied together for any price/quantity combination on a linear demand curve.

Comparing competition and monopoly

From this comparison comes one of the standard economic results, that is, monopoly tends to reduce supply and increase prices, compared to the social maximum, which is cost-reflective prices. In the example above it would be where price equalled LRMC, that is marginal cost pricing. The chart shows the relationship between competitive and monopoly supply. Clearly the price rises and the supply falls with monopoly, and the profit maximising profit could be said to be 'bad profit', albeit economic profit. In practice, other issues such as quality of product and consumer demand based on brand loyalty have to be taken into account in any analysis of whether excessive profits are being earned. To illustrate the combined relationships of cost and revenue curves it is useful to take the straight line demand curve example with that of constant returns to scale in supply. The two points A and B are useful, showing the points where $MC = MR$ and where $AC = AR$ respectively. Point C shows where $TR = TC$.

The importance of points A and B arises because they reflect the difference between competitive (P_cQ_c) and monopoly pricing (P_mQ_m) outcomes. We have first to be clear about supply and demand because we have to distinguish between that which can be supplied from competitive markets and a monopoly supplier. The average revenue (demand) curve reflects the valuation for each additional unit supplied, being the price that the marginal consumer would be willing to pay to buy it (as compared with the marginal revenue curve which combines the revenue effect of an extra unit sold with the loss of revenue from selling all of the previous units at the lower price necessary to sell the additional unit). What price consumers pay is, however, a result of what producers and suppliers decide to charge.



For a monopoly supplier, the demand curve reflects the demand for their product at various prices, and they would decide to maximise their profit by setting a price which makes $MR = MC$, which gives the maximum point of difference between the TR and TC curves for the quantity demanded at that price (ie, on the line vertically from A because the slopes of the TR

and TC curves are the same at that point). Consumers, however, would feel that they had lost out. For those buying, they would be paying a price significantly above the cost of production, and for those not buying they would see that they could have been supplied at a lower price, such as point B, where supply would equal demand and TR still exceed TC.

Introducing competition with a choice of suppliers means that no supplier can treat the overall demand curve as the demand curve for their particular product (unlike the monopolist). This makes it impossible for competitive suppliers to set monopoly prices because if one tried to it would get no buyers, and others would supply at a price nearer to marginal cost. In effect, the industry supply curve is the summation of all of the suppliers marginal cost curves, and the demand curve for the individual supplier has become almost perfectly elastic. Given this, then point B would be the equilibrium point for supply and demand, and price would be cost-reflective, reflecting the marginal cost of supply. For clear reasons, this point is considered to be a 'social optimum' compared to monopoly supply pricing. It reflects the fact that at this point consumer surplus is maximised, given the cost of supply.

Competition: oligopoly, duopoly, monopoly and 'natural' monopoly

Describing the market in operation means that a choice has to be made from a continuum of classifications. At one end there is 'perfect' competition, which is really a theoretical construct for analysis, because in practice there are many factors which make competition less than perfect. A baker selling bread in a high street can probably charge a little more than minimum cost because local customers would probably want to avoid travelling to the next high street in order to buy their bread. The next position is oligopoly, where there are only a relatively small number of suppliers. When this becomes only two, it becomes duopoly. The problem with these is that they can tend to form 'cartels' in which they collectively agree to set their prices and thereby extract a 'quasi' monopoly profit. In general, governments declare cartels to be anti-competitive, and regulate them. Cartels tend to be unstable because one firm might decide to take advantage of a sudden reduction in the cartel's agreed price in order to attract customers and attract a larger, long term, market share. Government regulators can try to facilitate such instability by offering not to penalise a cartel member who 'blows the whistle' and reveals the cartel to the regulators, who then proceed to punish the others. This could be termed 'incentive' regulation.

Finally there is the distinction between monopoly and 'natural' monopoly. A supplier which is just better than its competitors may become a monopoly, and this may not be a bad thing for customers who want to buy the products, but their protection in the long run against undue monopoly profit is that it is not a 'natural' monopoly. Competitors can learn and catch up, and will be particularly incentivised to do so if the monopolist simply seeks to take advantage of its current monopoly status by charging too high prices. A 'natural' monopoly occurs where the production process is such that one supplier can always supply at a lower cost than the sum of costs from multiple suppliers. Typically this is found in infrastructure industries, such as water, energy, transport and communications. Even here, though, technical change can undermine a natural monopoly. Competition is a dynamic process, and monopoly is not simply a fixed equilibrium.

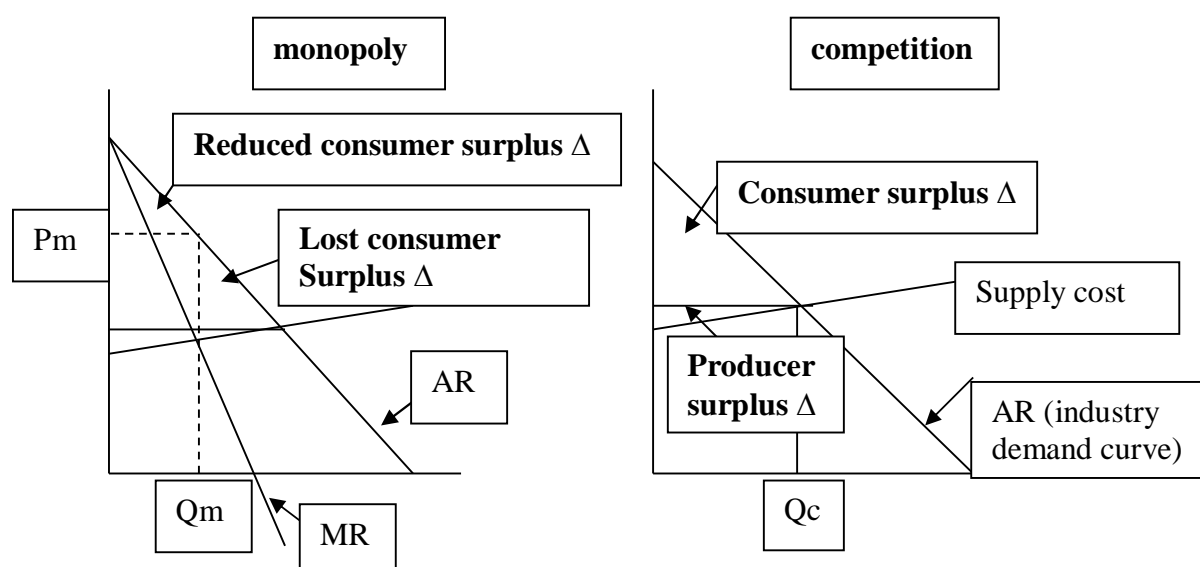
Consumer and producer surplus

Given the demand and supply curves, consumer surplus is the difference between what consumers would have been willing to pay for each item of the commodity being supplied and the market price they have to pay (ie, the area between the demand curve and the horizontal market price line). Producer surplus (or 'rent') is the difference between the market price the supplier receives and the cost they would have incurred to supply each unit. The relationship is usefully shown comparing the monopoly and competitive positions. The two charts compare

similar situations, but where the demand curve is the monopolist's demand curve in one, and the industry demand curve for the other. The supply curve is shown as a rising long run cost of supply. The charts show that the monopoly price is higher than the competitive market price, and that the monopoly supplier will supply less than the competitive market. It is important to note that as we move from a competitive to a monopoly position, some of the consumer surplus has been converted into profit for the monopolist (redistributed), and some lost completely due to the fall in quantity supplied (area ABC).

Given there is a real loss of consumer surplus from monopoly supply, the consensus is to consider monopoly profit as 'bad' profit (compared with extra profit from efficiency improvements), and a reduction in the social welfare maximum. Such analysis means that most governments have policies to promote competition and to regulate monopoly power.

Consumer and producer surplus



Types of Cost

Applied economics and accounting is much concerned with various types of cost, whether for making optimal decisions, or for simple reasons of effective accountability. These are in addition to the standard total, average and marginal costs that we have considered. For example, a distinction can be made between variable costs, which change with output, and fixed costs, which do not. A firm producing books for sale will have variable costs of materials used to produce the books, but the annual rent for the factory in which the books are produced might be fixed, irrespective of the output of books. Capital assets used in production, such as machines, can also be fixed costs in the short run. In the longer run they can become variable costs because the machine need not be replaced, or be sold second hand.

In addition to the distinction between fixed or variable, and short or long run, there is the important distinction between operating costs and capital costs. This division is important when we come to accounting because they are accounted for in different ways, for good reason. The key concept is that operating costs are those costs incurred in a particular accounting period, and which do not relate to a useable resource in subsequent accounting periods (eg, wages and salaries, heating and light). Capital assets provide services in the current period and can continue to do so in subsequent periods (eg, machines). Another important cost concept is 'opportunity' cost, which is the valuation of a resource in relation to the value that it would have

in its 'next best' available use. In deciding whether to produce or not, account has to be taken of the market value of the capital assets you currently have, as it might be best to sell the asset.

The three costs of business

The distinction between operating and capital costs can be brought together by listing the three costs of business, the three costs which must be covered by revenue if a profit is to be made. Some assets will wear out through use, resulting in a decline in value, called depreciation. The idea that depreciation reflects the fall in the value of an asset over time is a straightforward one, but in practice the way it is accounted for can be quite demanding, given the range of methods available. Depreciation of an asset could be treated as though the decline in value was equal each year. Divide the asset value by the life of the asset (C/L) and we have the 'straight line' depreciation charge for each year. The profile of depreciation charges over the life of an asset can be varied from straight line depreciation by 'accelerating' it (more is charged in earlier years, less in later years), or 'deferring' it (less in earlier years, more in later years). Depreciation charges could be directly related to the 'market value' of the asset, the annual depreciation being measured as the difference between the value of the asset at the beginning of the year and the end of the year. This is usefully termed 'economic' depreciation.

Other assets may not depreciate through use, such as land. In practice, accounting will be concerned with many types of cost, for example, depreciation might be offset by maintenance, or assets might be leased rather than bought. This will be shown in the section on accountancy, but here we want to link the idea of the three costs of business to the long run cost of supply and a constant annual sum. The three costs of business (and the cost of capital is looked at in some more detail in Appendix 1) are:

- operating costs
- cost of capital consumption (depreciation)
- cost of capital finance (the rate of return r^*)

A simple example is constant returns to scale, such as production which requires a machine, and that machine has a fixed capacity limit (say 1000 units per week). If demand increases to more than its capacity, you have to invest in another machine, and so on. If we have the constant annual capital cost related to each machine then we have the long run cost of supply for any level of supply. The average cost per unit of supply can also be calculated.

The method is to link the two capital costs of depreciation and capital finance together over the life of the capital asset in question using 'annuitisation'. The constant annual sum can be split into its depreciation and interest cost elements if required. The total capital cost can be added to the operating costs to give the total annual cost of supply. The same principles apply to people taking out a mortgage from a bank to buy a house, or to someone leasing an asset to use in their business. A constant annual sum is clear and straightforward, and the principal repayment and interest element can be separated out if necessary for tax purposes or financial reporting.

An example draws attention to two important aspects which arise from annuitisation. First, that it embeds a key financial reporting rule, which is that interest is earned 'on' the outstanding balance until the return 'of' the balance by way of principal repayment. Secondly, the profile of the repayment series will be compound, each years principal repayment increasing by $(1+r)$. The calculation therefore starts with the sum borrowed, on which interest is due for a year, before some of that sum is repaid, and the next year's interest is on the reduced sum to be repaid. The following example is for a loan of £100, repayable over three years and at an interest rate of 10% per annum. The sum of the discount factors at a 10% interest rate for three years is 2.487, therefore the constant annual charge would be $100/2.487 = 40.21$.

Year	Annual charge	Interest due	Principal repayment	Balance o/s
				100.00
1	40.21	$(100 \times 0.1) = 10$	$(40.21 - 10) = 30.21$	69.79
2	40.21	$(69.8 \times 0.1) = 6.98$	$(40.21 - 6.98) = 33.23$	36.56
3	40.21	$(36.6 \times 0.1) = 3.66$	$(40.21 - 3.66) = 36.55$	0.01
Total	120.63	20.64	99.99	0

The example shows that the repayments are increasing each year by 10% (ie, times 1.1) and the interest paid falls as the balance outstanding falls. The total sum paid is 120.63, being 100 for principal repayment and 20.63 interest charges. This is a lower interest charge than would be the case if the loan was only repaid at the end of the three years. The interest paid would then be $100 \times 1.1^3 - 100$ which is 33.1.

5. Cost-Benefit Analysis

Cost-benefit analysis is what it says it is - an analysis of the costs and benefits associated with a particular decision to be made. It is analysis which is suitable for both business and government decisions, but is most noticed in relation to government decisions on a wide range of things, whether road schemes, hospitals, the environment or regulation. A cost-benefit analysis will typically accompany public consultation, setting out the various options and focusing on that one which is judged to be best.

The important thing to recognise in cost-benefit analysis is that it is intended to be comprehensive and complete, covering all facets of costs and benefits. Its reputation suffers because some believe that cost-benefit analysis is only meant to take account of things which have a market price, and therefore important things are disregarded. It might be that some cost-benefit analyses are badly done, but it does not make the technique wrong in principle. It is helpful if the various aspects in a cost-benefit analysis can be given a value as that allows the costs and benefits to be added up and compared, and also gives a focus to the debate as to whether a fair value has been given. Take for example a road bypass scheme which has two potential routes, one which is longer and more expensive, and one which is shorter and cheaper, but happens to cross a heritage site. Clearly we might all feel that the extra cost of the first option is best, but even though we do not have a market in heritage sites, we would be giving an 'implied' valuation at least equal to the cost difference between the two options.

Risk is an important is in cost benefit analysis, as it is in business decisions. Government has to look for public support, however, where some actions might be costly. Speed controls and seat belt requirements are intended to reduce the risk of accidents or injury through accidents, and generally there is public support for such regulatory measures. Regulation can be more difficult where the risk is seen less certain or understood. If some people think that climate change is not caused by human activity, and climate change is not something which is a real problem now, then it makes it more difficult to impose restrictions now, which are costly, to receive a benefit sometime in the future. It makes the need for good evidence and rationality of argument in cost benefit analysis very important.

Studies into how to find a value for something in cost-benefit analysis are often undertaken. The benefit of investing in roads includes the saving in travel time. Estimates of the value of travel time might be found from seeing how many switch to a new route which includes a toll. If the new route is an hour faster than the old one, but costs £5, then that is a minimum value of their time. Consumer surplus calculations are also involved when investment in a new road increases the demand to travel. Valuation is important where the public value something but

don't use it. The rail network is a good example. A rail line may not attract sufficient fares to cover its costs, so should the government support it with a subsidy? Many who use cars are happy to know that there is a rail network operating if they need it. This comfort is something that they are willing to pay for, but there is no market to achieve that. For that reason, governments have to take account of their 'option' demand, and provide for that demand by keeping the railways open, financed from taxation.

Attempts to provide a valuation can always be controversial. Take for example the value of life used in transport cost-benefit analysis. If scarce financial resources have to be allocated between a large number of competing schemes, then it can be helpful for a particular scheme expected to save three lives over the next ten years to be able to say what benefit that represents in relation to its construction cost. If the government judges on our behalf that the value of a life is £1.5m in relation to public expenditure, then there is a total to compare with the construction cost. However, there will be those who say that such a cost-benefit analysis is illegitimate because you cannot value a human life. But if no account is taken of it the implication is that the value of life for public decisions is infinite, and clearly governments in general do not expend any amount of money to reduce some risk to life. Some valuation is implicit in the decision making process.

Cost-benefit analysis has therefore to be all-inclusive, but not arbitrarily constrained. It has to take account of social and environmental matters, and be open for debate. Only from this approach can decisions be trusted and be seen to be a legitimate decision taken in a balanced and fair way. For the government, cost-benefit analysis is really a form of 'political economy'.

Cost-benefit analysis, welfare and the regulatory state

To understand the role of cost-benefit analysis it is necessary to understand the context within it is used. Governments are empowered to do many things, but in essence they are concerned with issues arising from 'public goods' and 'public bads'. The four main categories are to control:

- * The abuse of monopoly and market power
- * Externalities (spillovers)
- * Provide public goods for which there is no market
- * Overcome inequitable distribution of wealth and opportunities

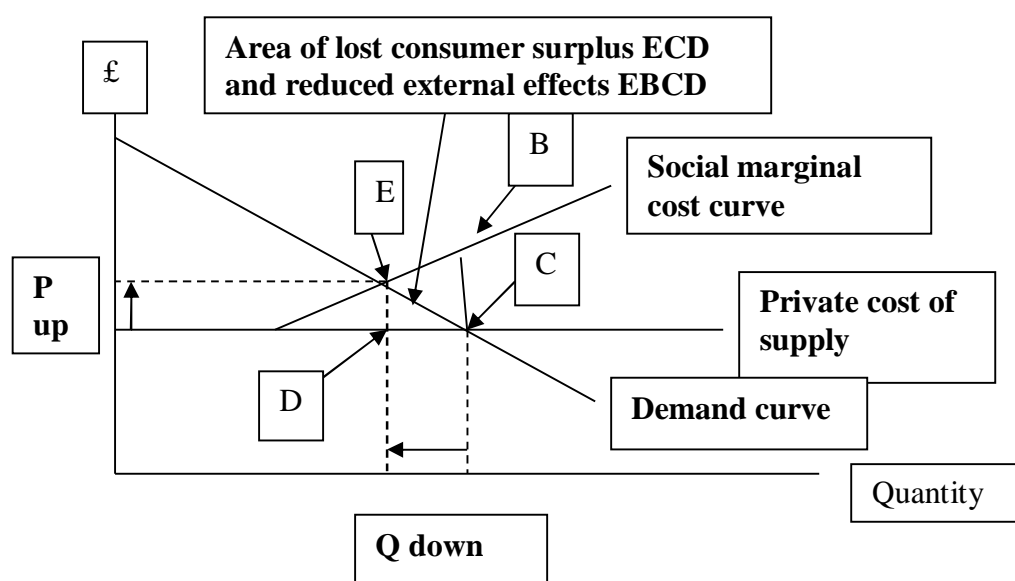
In fact all of these can be referred to as examples of 'market failure', which need government action to provide an effective remedy.

The abuse of monopoly power can be dealt with by having an effective competition policy and regulatory framework, such as a Competition Commission. Externalities are affects on third parties which result from the processes of supply and demand, for example, pollution from many sources. The policy incentive in relation to externalities is to find ways in which the costs of such externalities can be 'internalised' into the production and consumption decisions of those engaged in supply and demand. If emitting carbon is a cost because it causes adverse climate change, then the cost has to be included in demand and supply decisions. One approach might be 'command and control', which simply regulates by banning emissions above a certain level, another might be market based, where there is a price which has to be paid to emit carbon, and that price inevitably moves on to consumers who buy products which have required carbon emissions.

The idea of internalising an externality can be represented by a chart which incorporates not only the private marginal cost of supply but the cost of the externality as well, a 'social marginal cost curve'. In this case, the damage from the externality is assumed to start after a certain level of goods supplied, and then to rise for each one consumed thereafter. The point is to move

from point C, where demand equals supply based purely on the private cost of supply and unaffected by the cost of the externality, to point E where the social cost of the externality is taken into account. The trick is to internalise the cost of the externality into the supply price to consumers, so that the price goes up as shown and the quantity supplied goes down. The effect is that producers have to pay the cost of the externality to supply the goods, and consumers have to bear that cost when they buy the goods. Total welfare is increased because the loss of consumer surplus (area ECD) is offset by the gain through saved reduced external effects (area EBCD).

Internalising the externalities by pricing



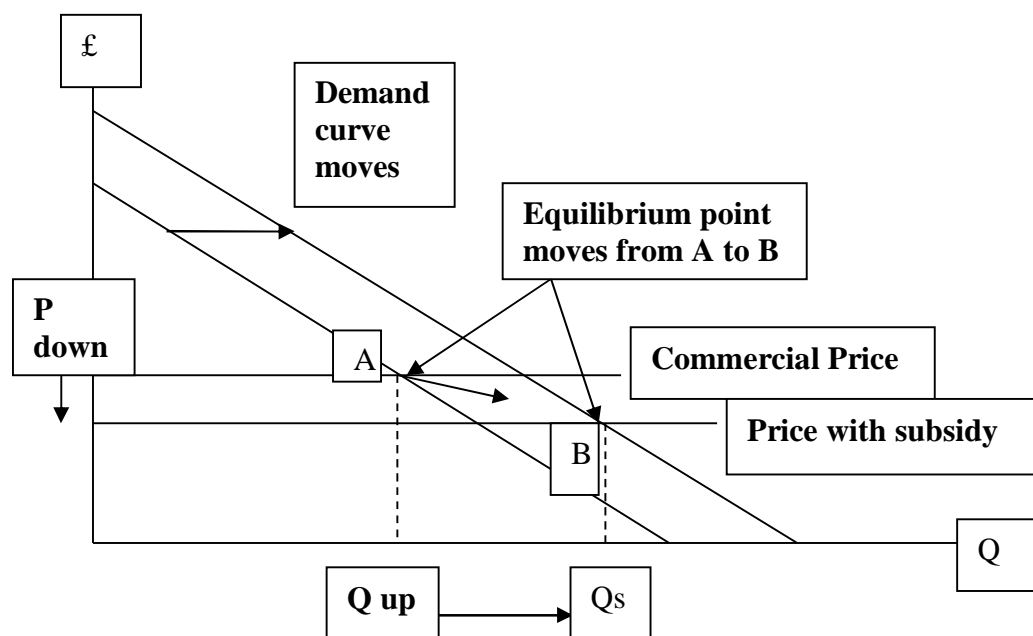
The opposite occurs where policy demands that something should be ‘subsidised’. Whereas pollution is bad for your health, vitamins can be good for your health, and a government might decide that the people should eat more fruit than it does now, and is willing to finance the cost of achieving it, given the expected benefit to both individuals and the cost of the national health service. Two approaches might be involved. First, a publicity campaign to try and encourage people to demand more fruit and move the demand curve, and secondly, to reduce the price of fruit to the consumer, perhaps by getting retailers to issue coupons to those who buy fruit, which can be used to purchase something else. This can be shown in a chart like the above, combining both approaches but where the price now goes down and the demand goes up, moving the equilibrium point from A to B.

The choice between subsidy or making the polluter pay arises many times when governments make regulatory decisions and this is a relevant issue when we cover the political economy of climate change.

Public goods arise where there would be no supply of those goods because the market could not prevent consumers from taking a ‘free ride’. National defence is a classic example because if it is provided for one it is provided for all, and no one can be excluded, so why should any one offer to pay. The market cannot respond to this, only the government can by forcing all to pay for it through taxation. Education is another example. As we all benefit from having an educated population, it is important to ensure that the state provides a standard system of education for all. This can be complemented by private, market-based education, but this is on

top. Public goods embrace the areas necessary to ensure that there is an equitable distribution of wealth and opportunities, and that essential needs are met. Health care needs to be available to all, and the less capable need to be given the opportunity to achieve. Regulation on businesses is introduced to achieve this.

Internalising public benefits



The concept of social welfare is therefore very important in considering how cost-benefit analysis should be used to inform government decisions. Some interpret this in a strictly 'utilitarian' way, saying that government should pursue the 'greatest good for the greatest number', in accordance with the philosophy of Jeremy Bentham. But what is optimal in this sense? Is it sufficient to say that the benefit for some should be sufficient to outweigh the costs to those who do not benefit? Or should it be a requirement that if there is to be a net gain, then it should be redistributed in such a way that the losers are compensated for their loss? Progressive taxation is clearly an important tool in achieving an equitable distribution of wealth. Cost-benefit analysis is a vital component of the business of the regulatory state and the practical operation of effective democracy. In some ways this is similar to private industries which report their performance by a 'balanced scorecard', including not just current profit data, but those things which may ensure that the profits are sustainable, including measures of customer care, investment in innovation, care for the community and the environment.

In effect, economic analysis covers all operations of the state, because all of its actions reflect the need to regulate some form of market failure, whether that requires regulation of markets and private activities, or provision of state services. Market failures can occur from abuse of monopoly power, externalities which arise because activities affect 'third parties', the interests of which are not taken into account when goods or services are supplied, the existence of public goods where markets do not supply them because a private supplier could not recover the costs and, finally, the existence of an unacceptable distribution of wealth or opportunities which would arise in an unregulated market.

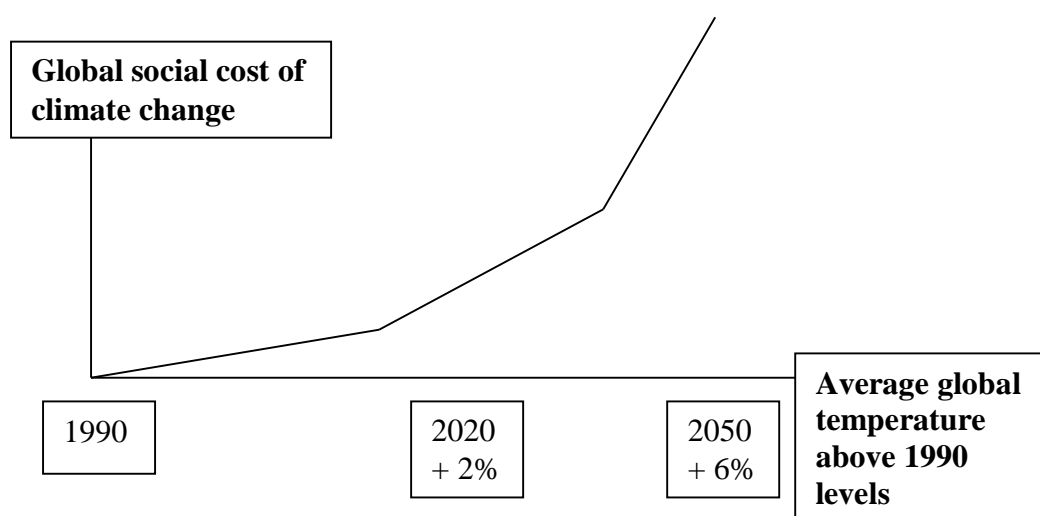
Applied Economics - Controlling Climate Change

1. 'Macro' political economy

The problem of controlling damaging climate change is a good example of 'macro' political economy, alongside international money markets and global poverty. It is a 'global' problem because emissions in one place affect the global climate, and therefore action requires a global response. There are, however, many reasons why appropriate international action may prove to be too difficult now, and prove to be too late if taken in the future.

The global climate eco-system is complex, and one problem is the distinction between climate and weather: climate is what you expect, and weather is what you get. The climate is the long run trend of the weather systems in place around the world. The average temperature of the world may be increasing but this might be hard to notice, or care about, if the regional weather system is generally cold and getting colder. Climate trends and regional weather fluctuations occur naturally, so the first problem is to identify the change in the climate and weather systems which are specifically caused by human activity. The forecast outcomes for climate and weather systems with 'human-induced' causation then have to be compared with the forecasts of what would have happened with no significant human-induced factors. Forecasts for such a complex system means that even if scientists can achieve a consensus, it may be difficult to persuade the public in general.

The debate is framed in terms of 'greenhouse' gases, so that if human activity has led to an increase in emissions of carbon dioxide (CO₂) into the atmosphere, this causes an increase in the average global temperature. The broad consensus is that if the global temperature increases by more than 2 degrees centigrade over the 1990 level, the resulting effects might be 'catastrophic' in terms of melting ice caps and rising sea levels, and dramatic changes in regional weather systems, as simply illustrated in the diagram. There is no shortage of debate about this potential catastrophe, given books such as **'Where on earth are we going'** by Jonathon Porritt (1990), **'Fragile Earth – Views of a changing world'** HarperCollins (2006), **'6 degrees – Our future on a hotter planet'** by Mark Lynas (2007), and **The Burning Question: We can't burn half the world's oil, coal and gas, So how do we quit?** by Mike Burners-Lee and Duncan Clark (2013).



However, it is not just carbon emissions that have to be taken into account. Methane from cattle in agriculture is a much more 'potent' greenhouse gas than CO₂ per ton, and so the measure has to be standardised to tons of CO₂ equivalent (tCO₂eq). In addition, it is not just

emissions to consider, but ‘absorptions’ too, as trees absorb CO₂. Cutting down a forest is therefore the same in principle as emitting CO₂.

Notwithstanding the complexities of the forecasts, any system to be put in place to control the ‘net’ omissions of tCO₂eq has its own complexities. The simplest approach might be to allocate the same proportionate reductions in tCO₂eq to each country, assuming that there was a global consensus on the maximum net global amount of tCO₂eq which can be emitted in any particular year, and on the forecast of the global tCO₂eq which would otherwise be emitted. This means the global maximum of tCO₂eq allowed divided by the \sum forecast tCO₂eq for each country = % reduction required by each country. Two problems are immediately evident, however. First, some countries would be required to make reductions when their forecast emissions are well below the average emissions and, secondly, the cost of reducing emissions to each country could not be taken into account. The outcome might not be cost-effective or equitable. To overcome this, an illustrative global approach can show the advantage of ‘carbon trading’.

2. An illustrative global approach using carbon trading

The following example is based on the idea that the international community (reflected in the United Nations) has both agreed to the global totals of allowed net tCO₂eq emissions each year and to setting up of the necessary institutional arrangements to operate an effective carbon trading scheme, given emitters and absorbers would have to be registered, and the outcomes audited. In addition, the registration scheme might only apply to the commercial and business sectors because it would be rather inefficient to include the domestic sector when the cost of their use of fossil fuels (eg, driving a car or using a gas boiler) can be focused on those who sell them the fossil fuels.

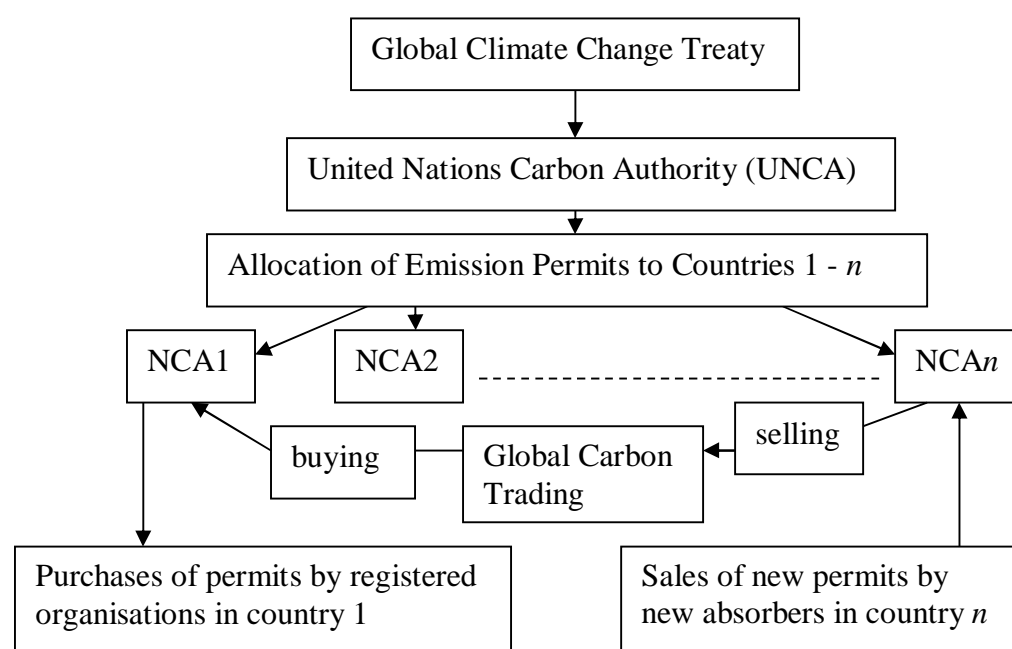
The three economic principles embedded in this example are externalities, the ‘polluter pays’ principle and the ‘precautionary’ principle. Net tCO₂eq emissions are a polluting externality which affects third parties without the polluter having to take account of the cost. The polluter pays principle simply requires the polluter to bear the cost, but accepting that that cost will be passed onto customers who buy the goods and services produced. Given this, the supply and demand for the goods and services will be balanced, including the ‘cost of carbon’. The precautionary principle simply requires action to be taken when the risk is ‘uncertain’ but the costs of not doing something if the outcome occurs vastly outweigh the costs which would be borne if in the event the outcome did not occur. In short, gamble on what you can afford, but insure against what you cannot afford!

The prime institutional requirements would be for the United Nations to establish the UN carbon authority (UNCA) and for each country to establish its national carbon authority (NCA). The role of the UNCA would be establish, in accordance with the treaty, the global sum of tons of CO₂eq which can be emitted each year, and to agree the profile of the global emissions sum over the years ahead (say to 2020, 2030 and 2050). Given the global sum, the UNCA would divide it by the global population to give the allowed tons of CO₂eq for each citizen. The UNCA would then allocate to each country its share of the global sum based on its population. A country with a large population would get a larger allocation than a country with a small population, but it would be equitable in that each gets the same amount per head.

The role of each country’s NCA would be to supply emission permits to registered organisations in their country who wish to apply for the permits to cover their forecast tCO₂eq emissions. The NCAs could trade among themselves in a global market to buy more permits where the

demand in a country exceeds its allocation from the UNCA, and to sell permits where the demand is less than their allocation. The result of trading would be to balance supply and demand and create a global price per ton of CO₂eq, and the NCA would sell its permits to applying organisations at that price. Where a large landowner intends to cut down an absorbing resource, then the landowner would have to apply for emission permits each year based on the increase in net tCO₂eq that the action of cutting down the forest incurred. Where a large landowner decided to provide an absorbing resource, then the landowner could offer to sell 'new' carbon permits to the NCA reflecting the addition to the global tons of CO₂eq which can now be allowed, leaving the net tCO₂eq sum the same. The landowner would receive the international price of carbon. The other role of each NCA would be to audit the operation of the scheme to ensure that the registered emitters are covering their emissions with permits, and to provide the national statistics to the UNCA.

The basic framework for such global carbon trading could be illustrated as follows:

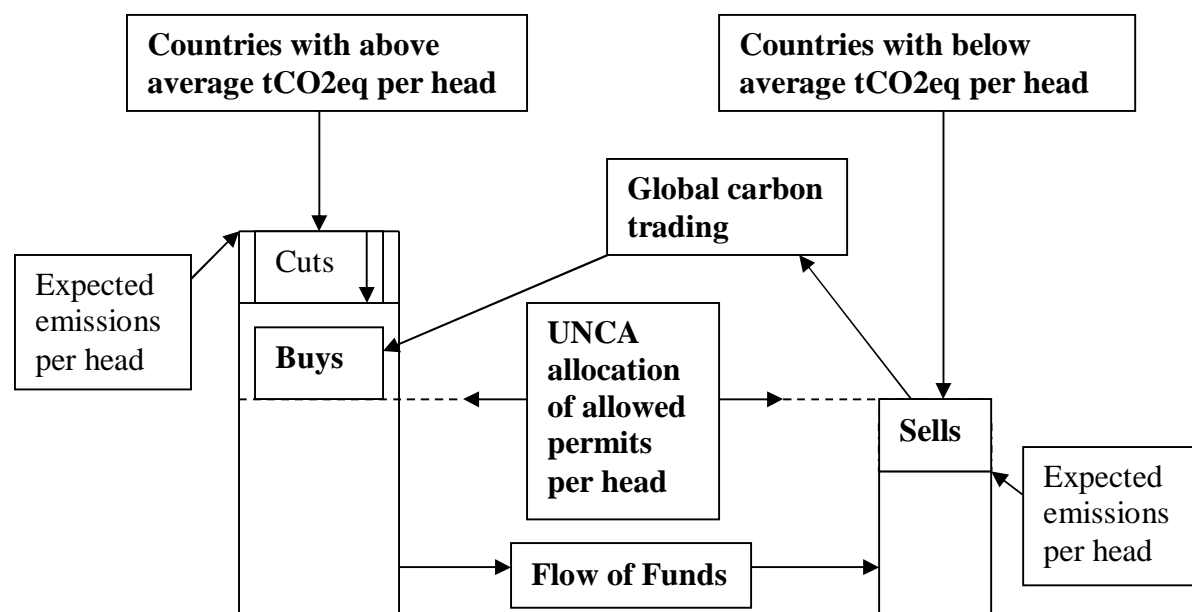


The equitable outcome

The financial consequences of such a trading system has two specific advantages. First, it minimises the global cost of meeting the maximum allowed tCO₂eq emissions targets. Secondly, it creates an international flow of funds from high emission countries to low emission countries. The reason for its cost-efficiency is that emission certificates will be purchased by registered organisations when the cost of reducing emissions is greater than the market price of carbon. Where the cost of reducing carbon emissions is less than the price of carbon, then it is best to reduce emissions. Given that the UNCA has issued only sufficient permits to meet the global maximum allowed emissions of tCO₂eq, all emitters must have permits to do so, and supply must equal demand, then the global trading market will produce a global price for a permit to emit a ton of CO₂eq. The excess of emissions between the allowed maximum and the global forecast for intended emissions must be met by cutting those emissions, but this will be done by those organisations where the cost of cutting emissions is less than the price of carbon.

The equitable flow of funds between countries can be illustrated by comparing countries where the emissions per head are greater than the global average with countries that have emissions per

head less than the average. In above average countries the excess must be covered by the NCA purchasing additional permits or by the registered organisations cutting their emissions. For below average countries, the NCA can sell the allocated permits not required in their country. The NCA in the above average countries will receive revenue from selling its allocation of permits from the UNCA plus any additional permits purchased from the global trading market, less the cost of buying the additional permits. The NCA in the below average countries will receive revenue from selling its surplus of permits from the UNCA to the global carbon market plus the revenue from selling the other permits to its registered emitters. Given that there is a correlation between high emitters and high GDP per head, and lower emitters with lower GDP per head, the flow of funds is an equitable way to tackle the control of carbon emissions. The relationships could be illustrated as follows:



The tax option

The global carbon trading approach can be compared with an alternative proposed by many, which is to tax the emissions. In one sense it looks the same because a tax to emit a ton of CO₂eq looks much the same as having to buy a permit to emit a ton of CO₂eq at the global cost of carbon. However, in practice it would be very different because it would not be globally cost-effective or equitable. The global price of carbon arises from market trading without governments having to decide what it should be, their role being to establish the global maximum of emission certificates. In the tax approach it would be a requirement on all governments to agree a global tax rate to emit a ton of CO₂eq and hope that it achieved the right outcome in terms of global emissions in total. Over time observations of whether the global tax rate had been too high or too low in terms of global emissions of CO₂eq could be made, and adjustments could be made to set it closer to the global price of carbon which would have emerged from carbon trading.

The bureaucratic difficulties of operating such a global tax approach are probably sufficient to make the carbon trading approach the right one, but there is also the problem that the tax approach misses the equitable outcome that there is a flow of funds from richer, higher emission countries to lower emission countries. With the global tax approach, the tax revenue simply goes to each country which collects the tax from its registered emitters (ie, emission certificates x tax rate) rather than where the NCA receives the UNCA allocation plus purchased certificates x global carbon price less purchased certificates x global carbon price (ie, each NCA receives

the UNCA allocation x global carbon price). The revenue from selling purchased certificates is simply transferred to the NCA selling them.

Inefficient subsidy

In terms of energy ‘mix’, the illustrative global carbon trading example would be a situation whereby the polluter pays (users of fossil fuels) and their rising prices will help create more of a balance with ‘renewable’ energy suppliers, which without carbon trading are often more expensive. The outcome is a better mix of fossil fuel and renewable energy, a lower demand for energy overall because the average price has risen due to fossil fuels having to pay for their emission certificates, and an increase in government revenue from the trading of permits through the NCAs.

The opposite is the case if it is decided to ‘subsidise’ the renewable energy suppliers rather than make the polluter pay through emission permits. In this case the mix of suppliers will improve because the cost to customers of renewable energy sources will fall, but the demand for energy overall will remain the same because it will simply be a transfer of customers from fossil fuel energy suppliers (FFESs) to renewable energy suppliers (RESs) at around the existing price levels. Government available revenue will fall as it has to finance the subsidy, therefore it will have to either raise taxes or reduce spending on other public services, unlike the carbon trading system, where taxation could be lowered or the surplus new revenue used to fund other new public spending.

The comparison of polluter pays carbon trading with subsidised renewable energy can be shown as follows, but this does not mean that subsidy is not appropriate in other situations. In general terms, for externalities affecting third parties, the polluter pays principle should be applied, but for ‘public goods’, such as cultural heritage and the arts, a public subsidy might be entirely appropriate. This includes such things as investing in research and development for renewable energy sources. Paying a subsidy to renewable energy providers could be justified however if it was considered a ‘public good’ in terms of reducing the risk to ‘energy security’ of having to rely on the import of fossil fuels from other countries. ‘Diversity’ of supply could also be considered to be a public good as insurance against the risk of future fluctuations in different sources of supply.

Comparing different options *outcomes*

*** No action on climate change**

FFES prices	a high mix	rising tCO ₂ eq
< RES prices	of FF generation	

*** Carbon trading system**

FFES prices	a balanced mix	less tCO ₂ eq	higher government
up ≈ RES prices	of FFES	through lower	net revenue
	and RES	demand and	available for lower
		better mix	taxes / more public
			services

*** Subsidy system**

FFES prices	a balanced mix	less tCO ₂ eq	lower government
≈ RES prices down	of FFES and RES	but only by	net revenue available
		better mix	requiring higher taxes /
			less public services

3. The political reality of climate change action

The illustrative example reflects an attempt to apply principles of political economy to controlling adverse climate change. The precautionary principle informs public policy, the polluter pays in order to reflect the externality they cause, demand and supply adjust to reflect the internalised cost of the externality (the incentives applying to customers, emitters and absorbers), the total cost is minimised (the global price of carbon will equal the marginal cost of tCO₂eq reductions, and noting that domestic users of fossil fuels, such as gas and petrol, can be exempt because they pay the necessary carbon price when they buy from the supplier of fossil fuels, who would need emission permits to sell the fuels), and governments net revenues are increased from of an equitable combination of flows of funds from polluters to NCAs in buying emission permits or from NCAs trading in the global market when they have surplus allocated emission permits from the UNCA. The system would have its administrative costs, setting up the NCAs and running the carbon trading system, and the audit process to regulate the system and ensure compliance, but any system would have significant administrative costs.

- The United Nations Kyoto Protocol

The strength of the illustrative example is reflected in developments which have taken place over recent decades. The United Nations have established the UN Convention on Climate Change (UNFCCC) with its Kyoto Protocol: an “international environmental treaty with the goal of achieving stabilisation of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system”. It was adopted in 1997 and came into force in 2005, with 191 states signing and ratifying it, but not the United States of America, and in 2011 Canada withdrew.

Four greenhouse gases were identified: carbon dioxide, methane, nitrous oxide and sulphur hexafluoride, and combined into CO₂ equivalents. The base year adopted was 1990 emissions, and the global maximum tons of CO₂eq to be allowed in any year are a percentage change from that base. The focus has been on ‘Annex 1’ countries. These 37 countries have to submit an annual report of their inventories of all anthropogenic greenhouse gas emissions (GHG), covering sources and removals from ‘sinks’. The institutional side is covered by each country having to establish a ‘designated national authority’ (DNA) which manages the GHG inventory. Practically all non-annex 1 countries also have a DNA in order to take advantage of the ‘Clean Development Mechanism (CDM)’. This is because the Protocol allows ‘flexible’ mechanisms for Annex 1 countries to meet their obligations by purchasing GHG emission reduction ‘credits’ from projects that would reduce or avoid GHG emissions in other countries (carbon offsets). International emissions trading has emerged, particularly in the European Union, with its emissions trading system (ETS), and in Japan, Australia, New Zealand, and the USA, with its Regional Greenhouse Gas Initiative.

The key elements of global targets, institutional frameworks and trading to incentivise cost-effectiveness are therefore in place, but it is evident that some countries have decided not to be involved, or have withdrawn, and the development of the system has a ‘hybrid’ sense of command and control with forms of carbon trading. The problem of ensuring that all countries are part of the global agreement, and its associated obligations, is an important one, because how are countries which want to be included to deal with countries who do not? Do the included countries have to take responsibility for even higher reductions in tCO₂eq in order to compensate for the reductions not taking place in the non-included countries? Or does it simply lead to a situation in which more countries opt out? Given the scientific outlook on the increase in the global temperature getting near the 2 degree centigrade threshold seen as a critical maximum, it is clear that effective, fully global treaties need to be in place before 2020.

Reasons for the impasse

One role of political economy analysis is to consider why the apparently straightforward solution to the problem of human caused climate change is proving so difficult and controversial to adopt. The reasons cover many aspects, and these fall into the following main categories:

- * The precautionary principle is not applied
- * The power of lobbying and trust in 'technical' solutions when necessary
- * The polluter pays principle is either under or over-emphasised
- * Attitudes and preferences are strongly influenced by a lack of trust, belief systems (particularly around the idea of 'profit' from carbon trading), and a 'historical' viewpoint on which countries are to blame for climate change.

To illustrate this, reference to some of the textbooks and reports over the past decades can be helpful, and provides an introductory bibliography.

The precautionary principle

Two books were published in 1973 and 1989 which provide good examples of the emerging concern over human caused climate change and with an emphasis on the needs of the developing countries: 1973 **Small is Beautiful – a Study of Economics as if People Mattered**, EF Schumacher and 1989 **Blueprint for a Green Economy**, D Pearce, A Markandya and E Barbier. However, by 1995 and 1998 two other books had been published which reflected a different perspective: 1995 **Small is Stupid – Blowing the Whistle on the Greens**, W Beckerman and 1998 **The Skeptical Environmentalist: Measuring the Real State of the World**, Bjorn Lomborg. The sceptics continue as we see from James Delingpole's book in 2011, **'Watermelons – the green movement's true colours'** which discusses how climate scientists are destroying the world.

Reading such books makes it clear that adopting a coherent global policy on climate change action is politically going to be very difficult. The public probably accept that there is a 'risk' of climate change catastrophe, but are uncertain as to how serious it is and whether something needs to be done now. It is always easy to postpone a difficult decision, particularly if there is a belief that the technology will be available if action needs to be taken in the future, and the ability to postpone is made easier if there is not consensus in the scientific community, or trust in the scientific community is undermined. The power of the self-interested lobbyists cannot be disregarded when it comes to the public, and politicians, deciding what action should be taken and supported.

Government policy

Each country has to consider whether its policies should proceed as part of a global approach or independently. There is a tendency to the latter, and this in itself makes it more difficult to succeed with a coherent global approach. Governments are concerned by public resistance to higher energy prices, so they tend to adopt the approach of providing subsidies. Governments welcome the principle of 'firm' government, and therefore tend to focus on 'command and control' systems rather than incentive-based systems. The UK government has taken such an approach by requiring an 80% reduction in tCO₂eq emissions by 2050. However, in terms of an equitable, cost-effective globally agreed system, there is nothing to suggest that the UK should cut by 80% rather than buy emission permits which fund other countries avoidance of growing CO₂eq emissions. The need to replace such a large reduction in fossil fuel energy generation has meant that governments are exposed to providing large subsidies to promote 'carbon capture and storage', biomass power, wind turbines (on shore and at sea), and the use of tidal and wave power. In addition Government's have to tackle the question of whether 'nuclear' power should be the preferred alternative to fossil fuels, notwithstanding the large lobby in favour of not allowing nuclear generation. Even if carbon capture and storage was technically feasible, it has its risks, as we can see from the existing 'methane' captured and

stored already in peat bogs and Siberia by the frozen land. It will only take a relatively small change in the climate to release the methane into the atmosphere.

Two particularly important effects of 'belief' systems are evident in terms of the political economy of global climate change action. First, there is often strong resistance to the idea of carbon trading because it implies that somehow people must be making a 'profit' out of it, and profit is considered to be a bad thing. Overlooked is the fact that its primary purpose is to minimise the cost of controlling climate change, and so the benefits of carbon trading are under-emphasised. Secondly, there can be a divide between the above average and below average CO₂eq emitting countries, particularly where the developing countries and some environmental campaigners believe that the problem has been caused by the developed countries over the last two hundred years, and that therefore they should be responsible for sorting it out. The polluter pays principle is over-emphasised but overlooked is the fact that a global trading system based on an equitable allocation per head of emission permits to each country would result in an equitable flow of funds to the developing countries, and gives an opportunity to avoid focusing on the past, but to focus on the position now and looking forward. The problem is particularly evident in terms of the incentives for absorbers of CO₂eq. Instead of recognition that those cutting down forests should pay for emission permits to cover the reduction in absorption capacity, it is assumed that landowners should be paid (ie, subsidised) not to cut them down.

However, whilst the difficulty of achieving a coherent global approach is evident, it is also clear that the pressure to achieve such an outcome is likely to succeed. The more debate there is, the more likely it is that the public will accept the argument that something should be done earlier rather than later, and that the cost should be borne. The debate was certainly increased in the USA in the 2012 presidential election following the devastating damage caused by 'Superstorm' Sandy on the east coast, when the popular mayor of New York, Michael Bloomberg, put his support behind Barack Obama, citing Republican challenger Mitt Romney's failure to back climate change measures. The range of books, reports and legislation is very large and the following gives some examples, with a focus on the report prepared for the UK government by Sir Nicholas Stern on the economics of climate change. To understand the political economy of climate change action is clearly a detailed area of study, although the main principles for policy action are reasonably clear:

1999 Pollution for Sale – Emissions Trading and Joint Implementation, eds S Sorrell and J Skea
2000 Energy – The Changing Climate, Royal Commission on Environmental Pollution, 22nd report, Cm4749

2005 Climate Change Policy, ed, D Helm

2005 The Economics of Climate Change, Select Committee on Economic Affairs, House of Lords HL Paper 12-I

2006 The Stern Review on the Economics of Climate Change, Sir Nicholas Stern

2006 The Stern Review: A Dual Critique, Part 1 The Science, Part 2 Economic Aspects, World Economics, vol 7, no. 4

2007 Regulating to Mitigate Climate Change – A Response to the Stern Review, Better Regulation Commission

2008 Climate Change and the Stern Review: The Implications for Treasury Policy, House of Commons Treasury Committee, HC231

In the UK there is now a Department for Energy and Climate Change (DECC) and a statutory requirement for five year carbon 'budgets'. However, the debate continues and it is questionable whether people's behaviour will change enough to allow the required emission

reductions to be achieved by 2020 and 2050. The headlines which continue in the press give a sense of the ‘shifting sands’: The Financial Times 5/2/2007 “Beware of Reckless Regulation on Climate Change”; The Guardian 12/3/2012 “UK wants Renewable Energy Target Scrapped”; The Guardian 22/2/2012 “Drax Scraps Plan for Two Biomass Power Stations, Citing Lack of Government Cash”; The Guardian 6/2/2012 “Concern Grows over Role of Scientists and Billionaire Backers urging Quick Climate Fix”; and The Guardian 12/3/2012 “The Toad and The Tortoise: Why Solar Power Schemes in the Desert are Feeling the Heat”.

To unravel the extremes of the debate about action on climate change will need both good analysis of its political economy, and the ability of political leadership to carry through the conclusions of that analysis via the United Nations and its Intergovernmental Panel on Climate Change (IPCC). There is a sense that climate change and cultural inertia are linked, and that debate on climate change policy has to tackle the balance between ‘an inconvenient truth’ and a ‘reassuring lie’. It could be that climate change as a global policy imperative is stalled by the fact that all our experience has been conditioned by ‘micro’ political economy (ie villages, towns, cities and countries), whereas this is our first major example of ‘macro’ political economy (notwithstanding such things as the two ‘world wars’ and the experience of dealing with the damage to the Ozone layer). It is ironic to think that it is not just ‘human-induced’ dangerous climate change that might have to be addressed. Natural climate change could become dangerous to human life styles and, unlike the past, where people could easily migrate, that is not the case now. Would there be any objections to our seeking to adjust ‘natural’ climate change in this case, or would we rely on the ‘Gaia hypothesis’ of a self-regulating natural world, developed by James Lovelock, to provide the answer?

‘Micro’ political economy

1. Questionable Outcomes – the power of ‘safety and security’

It is a very common experience for people to wonder why things turned out as they did. Was it because the outcome was pre-determined, irrespective of the analysis, the number of alternative options, and the strength of opinion against the project in question? Or was it because over time the circumstances changed and new information came to light? There is probably no right answer to questionable outcomes but in many cases there is a common element, the power of ‘safety and security’.

Two examples are very evident for the UK in 2012. First, the London Olympics. Headlines in the Guardian newspaper capture the issue, “Welcome to Fortress London – The Olympics will see the UK’s biggest mobilisation of military and security forces since the second world war”, and the current estimates for the security costs of the games have risen to £550m. Why is it that a ‘sporting event’ should make it so necessary to emphasise security over almost everything else? Secondly, the government’s attempts to reform the national health service (NHS). The NHS is a British institution which has given a feeling of security to the public ever since it was formed just after the second world war. To change it therefore summons up quite quickly a public mood that the policy is not very safe. Another paper headline captures the feeling, “Condition Still Critical – following a long and painful battle, the Health and Social Care Act finally made it onto the statute book. But after all that legislative surgery, the prognosis remains unfavourable for the NHS”. The first question asked is “So, after all that expenditure of political capital, what is the Health and Social Care Act actually for?”.

These two examples are at the national level, but the issues can be explored with an example of ‘micro’ political economy at the local level. The case in question is the infilling of the Combe Down stone mines on the outskirts of Bath Spa in Somerset. This is a project where some

concerns were raised in 1990 about the ‘stability’ of the mines, and the initial thoughts were that remedial work costing around £3-5m would be sufficient. By 2010 the mines had been completely infilled at a cost of some £180m. To understand how this happened it is necessary to look at the developments over the years 1990-2010.

Combe Down stone mines

The Combe Down stone mines were considered important because they provided the limestone in the 1700s and early 1800s to build ‘Georgian’ Bath, which is now a ‘World Heritage Site’. Bath is important because Georgian Bath was preceded by ‘Roman’ Bath. Romans lived in Bath until the 400s, taking advantage of the hot spring waters and building the Roman baths. The building of Georgian Bath was a coming together of Ralph Allen, who promoted the use of bath stone, the architects Wood the elder and younger, and Beau Nash, the socialite who made Bath a destination for aristocrats. Once Georgian Bath was built the stone mines for the most part ceased production from the 1850s, and the mines remained as a legacy of a unique development. Over the next 150 years Combe Down was built up, with houses built on the mine roofs, supported by the pillars. Occasionally an infilled mine ventilation or access shaft might subside, but there was no major house collapse, and for most people they lived unaware that there were any mines.

Around 1900 however, a minor road collapse drew the attention of the local council and surveys were done of the mines. By 1993 these surveys had revealed a number of things. First, some of the pillars supporting the mine roofs had become degraded, in part due to ‘pillar robbing’ of their stone, and in some places the roof thickness was less than two metres. The mines are home to bats which are protected by legislation and the EU Habitats Directive, particularly the greater and lesser horseshoe bats, and English Nature certified the mines as ‘sites of special scientific interest’. The historical and archaeological importance of the mines also became evident. These were well documented in the report ‘Historical and Archaeological Assessment of the Underground Quarries now known as Firs Mine and Byfield Mine’, by David Pollard (1994) and the report ‘Protection for Bats in Bath Stone Mines, Combe Down’, by Dr RE Stebbings (1994).

The position by this time was that some 15 houses were seen as potentially at risk, and the question was how best to approach this. First, the pillars could be strengthened, and there was experience of doing this because when the trams operated through Combe Down in the first part of the 20th century, the roads were reinforced where necessary. The next alternative was to fill in the area under the houses with cement, and at this point the first real problem arose because it was proposed for reasons of cheapness to mix cement with pulverised fuel ash (PFA), the waste from coal fired power stations. The government was interested in finding a use for the PFA slag heaps surrounding power stations and from their perspective, the more used the better. However, the problem is that PFA has strong toxic elements, and the Combe Down mines rest over a grade 1 water aquifer. The response was that a community group was formed called the ‘Forum’ which opposed the use of PFA. However, this opposition was countered by a group called the ‘Deadline’, which suggested that peoples’ lives were being put at risk by not proceeding to infill as much of the mines as possible.

Politicians act strongly in relation to concerns about risk to life, and so the local council and the government were becoming more of the opinion that the mines needed stabilising as far as possible. In 1994 Bath City Council published a report ‘The Way Forward’ supporting “stabilisation on a ‘once and for all’ treatment basis that would secure total funding from central government”. The option to carry out remedial work on certain pillars was also made more difficult when the government’s ‘mines inspectorate’ said after inspection that they were the

most dangerous mines they had seen. This meant for people working inside the mines, but that was soon taken to mean for the houses over the mines. Once this opinion took hold, it was not politically feasible to carry out a 'do-minimum' option and focus on the houses at 'high hazard' because if one house had additional support, why not all the others. The safety card had been played and the debate now turned to a solution which meant no PFA but infilling as much as possible.

However, cost-benefit analysis had a role to play here, and in 1997 a report was prepared for English Partnerships and Bath and North East Somerset Council by Ecotec research entitled 'A Cost Benefit Analysis of the Remediation Options for Combe Down Mines'. The conclusions make interesting reading, given how opinion had developed over the last seven years. The "central conclusion is that on the basis of current circumstances and the indicative judgements about the probabilities of future incidents the cost of stabilisation cannot be justified on the grounds of either urban regeneration criteria or planning blight. If stabilisation is undertaken the main rationale would be that the risk to health and safety is unacceptable". Setting out six options to be considered (from do-minimum to comprehensive infill) the report concludes "the cost-benefit analysis shows that, using economic criteria, the optimum scenario option is the do-minimum approach". Such a conclusion presented a real problem for both the central and local governments, and the Deadline lobbyists, because it is difficult to fund a project which is shown to be uneconomic. Nevertheless, the report recognised how the argument would be likely to proceed when it said "Cost-benefit analysis adopts an economic perspective; moral and political considerations are necessarily excluded from such an objective approach based solely on monetary values. These would need to be taken into consideration in reaching a decision on specific options. In particular the moral and political issues concerned with public safety remain. Our analysis can only attach indicative – not definitive – probabilities of risk given the understandable reluctance of experts to offer definitive figures".

Given Ecotec had identified the key drivers for a cost-benefit analysis would be "engineering costs and the probability of incidents resulting in injury and death", it is not surprising that the next stage would be to dramatically increase the risk of injury and death. The danger of a 'progressive' collapse of many pillars came into play. After 2000 the problem of PFA begun to disappear because the option was now to infill with 'foamed concrete', which because of its light mass would be less expensive overall. The other development was to suggest that planning blight was a real problem for property prices in Combe Down, and that properties would become uninsurable. With benefits going up and costs coming down the time had come for a new cost-benefit analysis. The protection of bats could be achieved by encouraging them to migrate to an adjacent mine (at some expense), and the historical legacy of the mines could be maintained by contracting Oxford Archaeological Unit to carry out a photographic survey of the mines and collecting some of the artifacts found in the mines. Once infilled, Combe Down could have an 'interpretation centre'.

As English Partnerships, the National Regeneration Agency, was now involved on behalf of the Office of the Deputy Prime Minister, the Land Stabilisation Programme (LSP) was launched in 1999. As LSP funding is available for the treatment of land liable to become derelict, neglected or unsightly as a result of collapse caused by disused mines, it was not surprising to find issued in 2002 The Derelict Land Clearance (Combe Down Stone Mines, Bath) Order. In 2004 English Partnerships contracted an Economic Appraisal by DTZ Pieda Consulting, and concluded "The appraisal supports Bath and North East Somerset Council's application to the Land stabilisation Fund for the proposed stabilisation of the mines, and proposed funding of £154.6m". In support of this grant of public money English Partnerships said "In Combe

Down, the risks to human health and safety are of paramount importance given the high population density of the affected area and the immediate risk associated with mines which have been considered by some experts as the largest, shallowest and most unstable of their kind in Europe”.

The project had now become full infill of 80-90% of the known mine void, which was called the combined infill stabilisation, and it is interesting to see how the cost-benefit for that and do-nothing compare. The cost of do-nothing was given a present value of health and safety costs at £194m and property costs of £83m. These costs essentially disappear with the full infill option. Clearly, to achieve that much health and safety costs from doing nothing means that regular mine collapses must be predicted (although there had been essentially none over the last 160 years) and the result of these would be regular loss of life. In English Partnership’s table Mine Collapse: Public Safety over Appraisal Lifetime (30 years) the total number of deaths predicted was 134, the number of serious injuries 336, and slight injuries 606. The first cost-benefit analysis in 1997 had predicted low numbers because the combination of limited areas of high hazard and a low probability of people regularly being present at any collapse, but now the cost-benefit analysis in 2004 predicts high numbers in each case.

Once started the infill project, for practical reasons, found that it was best to infill everything, so that the hope for an interpretation centre based around access to the remaining 10% of the mines had to be shelved. So what can one conclude about a project with started in 1990 focused on the do-minimum necessary at a cost of some £3m and ends up with a project in 2010 costing, once the environmental and other costs are added in, of some £200m? The amount of cement used represented the biggest cement order in the UK for many decades, which raised concerns about the amount of carbon emissions that must have resulted in, and the immediate thing noticed when the project was completed was an increase in ‘local ponding’ because the rain water used to distribute easily through the mines, now it doesn’t. One house has had to be pulled down and rebuilt because of effects caused by the infill.

However, because so many people had been encouraged to believe that their lives were in danger, they are happy that the mines have been infilled. If the bats find that they have been troubled, it is not likely to be a concern of many. The local council, being risk adverse, is pleased to have had a major project, funded by the government, and can now promote the project as an example of engineering excellence. For the few who reflect that more has been lost than gained, and an opportunity to develop a cost-effective inspection and remediation approach over time lost, the best thing is to recognise how such questionable outcomes occur in practice. A sufficiently strong lobby based on the safety card, given the institutional support of public bodies, and complemented by the interests of civil engineering firms, is very likely to result in the outcome seen in Combe Down.

Accountancy

- Financial reporting, management accounting and corporate finance

Financial accounting is the bedrock of financial reporting, which is most notably concerned with the annual accounts of companies and organisations. Such accounts are vital to the effective accountability of reporting organisations, whether it be companies to their shareholders, government to the public, or to people who are particularly interested in the reports, such as environmental campaigners, or political commentators.

1. Classifying accounting information

The accounts are drawn up to represent different elements of financial information, related to a specific period of time. Here we are concerned with annual accounts, but in practice, particularly for stock markets, quarterly accounts of companies may be important published information affecting share prices and performance evaluation. Cash flow statements are one important account, as it is important for confidence that a company's cash position is known. A company with insufficient cash may become bankrupt, even though its long term profitability might be sustainable. An organisation has to manage its cash flow effectively.

The other two main accounts are the balance sheet and the profit and loss account (P&L a/c). The balance sheet defines the position at a point of time (eg, balance sheet as at 31st March 2010), whilst the P&L a/c reflects the transactions over a period of time (eg, for the year 1st April – 31st March). The two accounts are normally linked as financial reporting integrates the opening balance sheet for the year with the profit and loss account to give the closing balance sheet for the year (which becomes the opening balance sheet of the next). This distinction between the main accounts reflects the classification of accounting information to be used in such accounts, and the most important classification is as follows:

- ***Assets and Liabilities:** Balance sheets

- ***Revenues and Expenses:** Profit and loss accounts

- ***Receipts and Payments:** Cash flow statements

The various classifications are very important, and mistakes are often made in this area. Payments are made in real time, but they do not necessarily represent an expense for the accounting year in question.

The three accounts are therefore linked 'dynamically' over time. The profit and loss account brings together the revenues and expenses attributable to each accounting period, in this case a year. The most important concept is 'matching', so that revenues properly attach to the expenses (ie, costs) that have been incurred in earning that revenue for the year. In this way, the annual profit or loss can be properly defined. The account will classify expenses in specific ways. Wages and salaries and energy use to heat and light factories and to run the machines are a clear and understandable form of expense, but there are other important expenses that need to be included. Depreciation charges reflect the using up of capital assets in production, and other provisions look into the future to provide funding for such expenses as decommissioning. Without such charges, the matching principle would not be properly applied, and accounting profit would be overstated. The profit and loss account each year is closed, and the remaining balance transferred to the balance sheet as a revenue reserve.

The balance sheet brings together the assets of the organisation and matches them with the liabilities it has to its owners and other long term funders. Assets can be 'tangible' fixed assets, such as buildings and factories, or 'intangible' fixed assets, such as the value of a brand or patent. Capital assets are referred to as 'fixed' because they are intended to be fixed in the business for the longer term. Current assets include such things as the value of stocks of

production and debtors, money owed to the organisation by consumers who have purchased on credit. Current liabilities include short term loans from creditors, or purchases made by the organisation using credit.

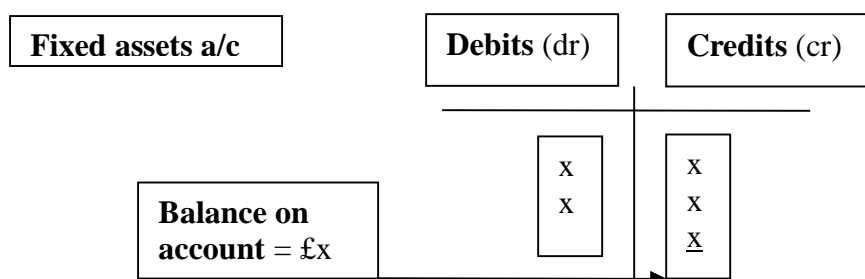
An important part of the balance sheet relates to the liabilities of the organisation to its owners. The owners have invested in share capital, which does not need to be repaid, and their return is by way of annual dividends paid on the shares, and by the market price of the shares, given that the shareholder can sell the shares. The profit or loss earned each year is therefore the shareholders money. Where it is not all paid out in dividends, then the remaining balance becomes a revenue reserve, being part of shareholders funds. Given that revenues and expenses are not the same as receipts and payments, this means that the link between the profit and loss account and the cash flow statement has to have an adjustment to convert one into the other, moving from 'profit from operations' to 'cash from operations'.

2. The 'double entry' system

The double entry system allows each transaction to be treated as two entries into the relevant accounts. At the end of the year, all of the accounts are added together in the balance sheet, and because all of the accounts are managed by the double entry system, the total assets will equal the total liabilities. Shareholders funds are a liability of the company, but are not usually referred to as such. Emphasis is placed on liabilities arising from long term loans made by banks. Assets less long term loans equals 'net assets', which is the same as 'shareholders funds'. The net assets, or 'shareholders funds', when added to the long term loans, equal 'capital employed'.

Two types of account arise. First, revenue and expense accounts. These are opened afresh at the start of an accounting period and closed at the end of the period, the final balance being transferred to the profit and loss account. Secondly, asset and liability accounts. These are continuing accounts and the balances on these accounts are shown in the balance sheet. The double entry system works because each account is deemed to have two sides (like the pages of an open book). One side is for debits (dr) and the other for credits (cr). In bookkeeping the debits may be on the left and the credits on the right. If an accounting system is being run on a computer, it works well because one can be a plus, and the other a minus. If accounts are being written out on a page, then the separation is achieved by a T diagram, with credits and debits being written on either side of the T, as shown in the chart.

The T diagram for double entry records



To examine particular types of transaction, and how they affect the various main accounts, it is best to use the two line approach, one for the debit and one for the credit. The account which is being debited or credited can be identified with its name. A few examples can illustrate this, remembering that debits represent 'assets or expenses', and credits represent 'liabilities or revenues', depending on the account being used.

The Double Entry system: example figures in £000s

Balance on **cash a/c**

First, a company raises money by selling share capital to investors:

Dr	cash a/c	£x	50	
Cr	share capital a/c	£x	50	50

Second, the company buys a fixed asset for cash

Dr	fixed asset a/c	£y	30	
Cr	cash a/c	£y	30	20

The company's cash balance is now £x-y so it decides to buy more fixed assets funded by a loan from the bank

Dr	cash a/c	£z	40	
Cr	loan a/c	£z	40	
Dr	fixed asset account	£m	20	
Cr	cash a/c	£m	20	40

The company employs someone to sell its products and pays them wages

Dr	wages a/c	£w	15	
Cr	cash a/c	£w	15	25

It sells its products, some for cash and some on credit

Dr	cash a/c	£r	25	
Cr	sales revenue a/c	£r	25	
Dr	debtors a/c	£d	15	
Cr	sales revenue a/c	£d	15	50

The company's machinery has a ten year life, and makes a depreciation charge

Dr	depreciation expense a/c	£c	5	
Cr	depreciation provision a/c	£c	5	50

This set of double entry transactions is sufficient to illustrate the relationship between the three main accounts. It is important to note that some of the double entries involve cash movements, thereby affecting the cash balance, but others do not, being accounting entries necessary to determine revenues and expenses. The profit and loss account brings together the revenue credited to the sales revenue account (£r+d) and the expenses debited in the wages and depreciation expense accounts (£c+w). The balance on the profit and loss account is £r + d – c – w, and would be a revenue reserve in shareholders funds, unless some of it was paid out in dividends. The balance sheet after these transactions would be:

Assets:			
Fixed assets	£y + m		50
Debtors	£d		15
Cash	£x – y + z – m – w + r		<u>50</u>
Total assets			<u>115</u>
Liabilities:			
Share capital	£x		50
Revenue reserve	£r + d – c – w		20
Long term loan	£z		40
Depreciation provision	£c		<u>5</u>
Total liabilities			<u>115</u>

and the total assets would equal the total liabilities. In presenting the balance sheet, the depreciation provision is taken over to the other side and subtracted from the fixed assets, giving the net book value (NBV) of the fixed assets (ie, $\text{£y} + \text{m} - \text{c}$). As each year's depreciation expense goes to the profit and loss account, the provision for depreciation increases until at the end of the fixed asset's life it equals the purchase cost, and it can be 'written out' of the books. For example this would be:

Writing out fully depreciated assets

dr provision for depreciation account	£80	
cr writing out a/c	£80	
dr writing out a/c		£80
cr fixed asset account		£80

and we can see that the fixed asset and provision for depreciation accounts have been reduced by the same amount and the writing out account has a zero balance. The debtors are written out of the balance sheet when they repay their debt (ie, dr cash a/c £15; cr debtors a/c £15).

For the cash flow statement, adjustments are made to convert the profit or loss balance to cash from operations, given some of the expenses in the P&L a/c do not involve any movement in cash, such as the depreciation double entry. Accounting for depreciation is a good example of the double entry system and the relationship between the three main accounts. A depreciation expense (charge) in the profit and loss account each year; an accumulating provision for depreciation of an asset in the balance sheet; and an 'add back' of the depreciation charge in the cash flow statement. The accumulating depreciation provision shows how money is being retained in the business to finance the replacement of capital assets, an idea called 'capital maintenance'. The money is retained in the business because the depreciation charge in the profit and loss account reduces the level of profit which would be available to distribute in dividends if there was no such charge.

3. The Main Accounts

The generic forms of the three main financial accounts are straightforward, as follows. In practice, the financial statements and annual reports of large companies are often complex and extremely detailed. Much information and analysis will not necessarily be found 'on the face' of the main accounts, but in the notes to the accounts. Nevertheless, the general idea of the three accounts is the same for the local greengrocer as it is for the largest companies, and the double entry system is the same for all.

The profit and loss account (income statement)

revenues and expenses for the period (dates)

This account goes through various types of profit before ending in retained profit, the sum which goes into shareholders revenue reserves. The cost of sales refers to all of the costs which are incurred up to the production of the finished goods. For a retailer, the cost of sales relates to the purchase of finished goods (eg, a shoe retailer buying shoes from a manufacturer). For a manufacturer, it is the cost of finished goods, for which there will be a separate account, the manufacturing account, which comes before the profit and loss account. As sales for the year may be drawing from existing stocks of finished goods, or some cost of finished goods for the year being costs relate to the finished goods being stored for sale next year, the cost of sales is calculated from the value of the opening stock for the year, plus purchases (cost of finished goods), less the value of the closing stock for the year.

The general format for the profit and loss account is:

The Profit and Loss Account		Example figures in £000	
	£		£
Revenue (turnover, income)	x		300
less cost of sales	<u>x</u>		<u>120</u>
equals gross profit	x		180
less other business expenses			
eg, wages and salaries	x	55	
energy use	x	22	
depreciation	<u>x</u>	<u>18</u>	
	<u>x</u>		<u>95</u>
equals operating profit (EBIT)	x		85
less interest paid(or received)	<u>x</u>		<u>28</u>
equals net profit (income before tax)	x		57
less taxation charge for the year	<u>x</u>		<u>15</u>
equals shareholders earnings	x		42
less dividends declared for the year	<u>x</u>		<u>18</u>
equals retained profit	x		24

The cost of sales analysis would be as follows, noting that purchases reflect purchases of finished goods by retailers from manufacturers, and the cost of finished goods reflects the manufacturing account for producers of finished goods:

	£
opening stock of finished goods	a
<i>plus</i> purchases or cost of finished goods	<u>b</u>
<i>equals</i> available for sale	c
<i>less</i> closing stock of finished goods	<u>d</u>
<i>equals</i> cost of sales	s

The manufacturing account precedes the profit and loss account with the cost of finished goods for the year. The basic format of the account is:

	£
Cost of materials used in production	x
<i>plus</i> factory expenses (labour, rent, power etc)	x
<i>plus/minus</i> the change in work in progress (WIP)	<u>x</u>
<i>equals</i> the cost of finished goods for the year	x

Provision for future liabilities is an important aspect of accounting. The 'joint cost' of an initial investment and the subsequent decommissioning cost is a good example, as both are, economically, part of the investment base, on which net revenues (ie, revenue less operating expenditures (opex) provide a return. The following example illustrates one method of achieving 'recognition' of the future decommissioning liability in the balance sheet, and 'matching' of revenues and expenses in the profit and loss a/c. Given a cost of capital of 10%, an initial investment cost of 400, and a decommissioning cost of 100 at the end of year 4 we would have:

$$\text{*Present value of decommissioning (PVD)} = 100(\text{df}_4) = 68.3$$

$$\text{*Annuitized value of PVD} = \text{PVD} / \sum \text{dfs} = 68.3 / 3.170 = \text{annual charge } 21.546\text{pa}$$

The total annual charge (expense) to the profit and loss account would be the annual charge of 21.546 plus the interest payable to the decommissioning fund (DFP) each year on the

accumulating fund balance. The balance on the DFP will thereby accumulate with interest to equal the future value of the decommissioning liability when it falls due at the end of year 4. The DFP a/c would look as follows:

Year	1	2	3	4
Opening balance on DFP	0	21.546	45.247	71.318
<i>plus</i> annual charge	21.546	21.546	21.546	21.546
<i>plus</i> interest on balance	0	2.155	4.525	7.132
<i>equals</i> total P/L a/c charge	21.546	23.701	26.071	28.678
Closing balance on DFP	21.546	45.247	71.318	99.678

At the end of year 4 the balance sheet will have a DFP a/c balance sufficient to meet the decommissioning liability and the profit and loss account will have spread that expense across the four years, thereby matching revenues and expenses. The expense charge over the four years will have reduced accounted profit and allowed the cash necessary to pay for the decommissioning to be built up. The double entry when the decommissioning cost comes to be paid at the end of year four would be:

Dr	decommissioning fund provision a/c	100
Cr	cash a/c	100

Operating profit is important, being earnings before interest and tax, and is used with a balance sheet number, capital employed, to give an important performance ratio, return on capital employed (ROCE). The net profit is really the end of the profit and loss account before it converts into the 'appropriation' part of the account. This is because taxes and dividend payments are judged to be an appropriation from profit.

The balance sheet

assets and liabilities as at (date)

The general format of the balance sheet is as follows, in the normal vertical format, rather than the horizontal format which keeps the debit balances on one side and the credit balances on the other, in the same way as a T diagram. Horizontal is a useful presentation for learning because it focuses well on how the double entry system balances assets with liabilities.

Most of the balance sheet items are self-explanatory, but some explanation is useful about working capital. Debtors are termed net debtors because account has to be taken of provisions made for 'doubtful debts'. If you know that some percent of your sales on credit may prove to be bad debts, then provision should be made for that percentage immediately otherwise your profit is overstated. Therefore, debtors less provision for doubtful debts equals net debtors. If you know that a debtor will not be paying then that debtor should be written off as an expense in the profit and loss account. Stock is made up of the cost of finished goods in the warehouse, work in progress and materials in store. The stock valuation is important to determining the cost of sales in the profit and loss account, and because stock might be purchased at different prices over time, there are various ways to calculate the stock value. It might be based on the 'last in-first out' (LIFO) method, the 'first in-first out' method (FIFO) method, or the 'average cost of stock' method. The advantage with the FIFO method is that it yields a fair current value of the stock holding shown in the balance sheet. Accrued expenses are expenses which need to be matched with revenues for the financial year; they are owed but have yet to be paid (eg, unpaid wages). Prepayments are when payment has been in advance for something which is to be accounted for next year (eg, an electricity bill might be paid in advance but one quarter of it relates to next year's accounts).

The Balance Sheet

		Example figures in £000			
		£	£	£	£
	fixed assets(tangible and intangible)	x		400	
less	provision for depreciation	<u>x</u>		<u>180</u>	
equals	net book value (fixed assets)		x		220
plus	other investments		<u>x</u>		<u>40</u>
equals	total capital assets		x		260
plus	current assets				
	stock (inventory)	x		60	
	debtors (a/cs receivable)	x		30	
	prepayments (of expenses)	x		15	
	cash in hand (at bank)	<u>x</u>		<u>8</u>	
		x		113	
less	current liabilities				
	trade creditors (a/cs payable)	x		20	
	accrued expenses	x		5	
	dividends payable	x		35	
	taxes payable	x		15	
	short term debt repayable	x		18	
	cash overdrawn (at bank)	<u>x</u>		<u>0</u>	
		<u>x</u>		<u>93</u>	
equals	net working capital		x		20
plus	other provisions		<u>x</u>		<u>40</u>
equals	capital employed		x		320
less	long term loans/bonds (debt)		<u>x</u>		<u>80</u>
equals	net assets		<u>x</u>		<u>240</u>
Financed by:					
	share capital (equity)	x		150	
	revenue reserves	x		60	
	capital reserves	<u>x</u>		<u>30</u>	
	shareholders funds		<u>x</u>		<u>240</u>

The cash flow statement

receipts and payments for the period (dates)

The link with the profit and loss account can best be shown by starting the cash flow statement with operating profit for the year, and the link with the balance sheet by ending with the change in the cash balance for the year, followed by the reconciliation with the cash balance shown in the opening and closing balance sheets for the year. The opening cash balance plus the change in the cash balance equals the closing cash balance.

The cash flow statement involves some adjustments to arrive at the right change in the cash balance from the operating profit in the P&La/c. Profit (or loss) on the sale of fixed assets has to be subtracted from (or added to) the operating profit at the start. This is because the receipts from the sale of fixed assets are accounted for later in the cash flow statement, and if the profit or loss was not removed, there would be double counting. This is because receipts from sale – net book value of the asset = profit or loss on sale. This means receipts = NBV + (or –) profit (or loss) on sale. Since the cash flow statement has a separate line for receipts from sale, then the profit or loss element is clearly not needed. That is for the profit and loss account.

The Cash Flow Statement

		example figures in £000s	
		£	£
	operating profit	x	120
less (or add)	profit (loss) on sale of fixed assets	<u>x</u>	<u>14</u>
equals	operating profit before profit or loss on the sale of fixed assets	x	106
add back	depreciation expense	x	35
plus/minus	accruals in operating profit	<u>x</u>	<u>(13)</u>
		x	<u>22</u>
equals	cash from operations	x	128
less	interest payments (receipts)	x	11
	dividend payments	x	33
	tax payments (credits)	<u>x</u>	<u>27</u>
		<u>x</u>	<u>71</u>
equals	cash available for investment	x	57
less	capital payments (capex)	x	108
plus	receipts from the sale of fixed assets	x	<u>22</u>
		<u>x</u>	<u>86</u>
equals	change in cash before financing	x	(29)
plus	new equity capital raised	x	50
plus/minus	change in net borrowing	x	<u>(19)</u>
		<u>x</u>	<u>31</u>
equals	change in cash balance for the year	x	+2

This change in the cash balance can be reconciled with the opening and closing cash balances in the balance sheet as follows:

		£	£
	opening cash balance	x	17
plus (minus)	change in the cash balance	<u>x</u>	<u>2</u>
equals	closing cash balance	x	19

Since depreciation is a double entry not involving the movement of cash, then it has to be added back because the depreciation expense has reduced the accounted profit but not affected the cash. Similarly, accruals have to be adjusted for. For example, the revenue from sales in the profit and loss account is made up of the cash sales plus the sales on credit, which may fall due for payment in the next accounting period. Hence a revenue change is not necessarily equal to the cash change. The adjustment ('unwinding' the accruals) is linked to the balance sheet, comparing the opening and closing balances for the financial year or period in question, and the rules are that:

- * An increase in an asset, or a reduction in a liability = a reduction in cash
- * A reduction in an asset, or an increase in a liability = an increase in cash

An example of such an accrual adjustment would be:

		£
	revenue in operating profit	1000
less	use of cash: increase in debtors	<u>200</u>
equals	cash from sales	800

A full worked example for preparing the three accounts follows.

4. Accounting – preparing a balance sheet, P&La/c and cash flow statement

The following example is for a manufacturer, Mr A Counts, who has been in business for two years. He therefore has an 'opening' balance sheet for year three, and its main features are as follows:

Balance sheet at 31.3.2011 (£000s)

Land	20	Share capital	50
Other assets (NBV)	60	Retained profits	20
Stock (materials)	20	Long term loan	25
Debtors	<u>15</u>	Creditors	10
		Overdraft at bank	<u>10</u>
Total	115		115

Mr Counts has financed his company from a balance of debt and equity, and has retained some profits, rather than paying dividends, as well as having an overdraft at the bank. The interest on the long term loan is 10% pa. Stocks of materials are made up of 50 'units' at an acquisition cost of £400 each, and each unit of sale requires one unit of stock. The stock is valued on the Last in First out (LIFO) method. The other fixed assets have a five year life, having been purchased at the beginning of year 1, and are depreciated on a straight line basis.

The accounts for year three can be prepared given the following information on transactions in the year:

- purchase of materials – 100 units at £500 each
- the stock of materials at the end of the year is 40 units, with no work in progress or finished goods for sale
- £3000 of existing debtors proved to be uncollectable (bad debts), the rest paid and £10,000 of new trade debtors were outstanding at the end of the year
- Given the bad debts, a provision for doubtful debts is set up at £2000
- Trade creditors were reduced by half during the year
- Wages of £22,000 were paid in the year with one month's outstanding
- Finished goods were sold at £1050 per unit
- Production expenses were £4500, which includes a 'prepayment' of £500 on an electricity bill taken in order to get a special discount
- Additional land costing £10,000 was bought during the year.
- administrative and marketing expenses came to £11,000

The manufacturing account

This has to take account of the valuation of the closing stock by the LIFO method, the depreciation charge for the year, the prepayment and the wages owing. The appropriate calculations are:

* **Stock used:** opening stock 50 plus purchases 100 less closing stock 40 = 110

* **Depreciation:** NBV at end of year 2 = 60 therefore acquisition cost of fixed assets with a five year life was $60 / 0.6 = 100$. The straight line annual depreciation charge is therefore $100 \div 5 = 20$ pa

* **Wages:** £22,000 for 11 months means £2000 owing ($22,000 \div 11 = 2000$)

Manufacturing account for year 3 (£000s)

	Opening stock – materials 50 x £400	20
<i>plus</i>	Purchases of materials 100 x £500	<u>50</u>
<i>equals</i>	Available materials	70
<i>less</i>	Closing stock (LIFO) 40 x £400	<u>16</u>
<i>equals</i>	Cost of materials used in production	54
<i>plus</i>	Production wages (£22,000 + £2000)	24
	Other production expenses (£4500 - £500)	4
	Depreciation charge	<u>20</u>
<i>equals</i>	Manufacturing cost of finished goods (110 units)	102

The Profit and Loss account

The profit and loss account can now follow, requiring in this case no specific, preparatory calculations:

Profit and loss account for year 3 (£000s)

	Sales revenue 110 x £1050	115.5
<i>less</i>	Cost of sales (manufacturing)	<u>102</u>
<i>equals</i>	Gross profit	13.5
<i>less</i>	Debtors written off	3
	Provision for doubtful debts	2
	Administration and marketing expenses	<u>11</u>
<i>equals</i>	Net operating profit (or loss) before interest	(2.5)
<i>less</i>	Interest due	<u>2.5</u>
<i>equals</i>	Net profit (or loss) after interest	(5)

Having made a loss of £5000 in year 3 Mr Counts has to prepare the balance sheet and consider whether he has enough cash to pay a dividend. Dividends are shown in the ‘appropriation’ account which follows net profit after interest, along with any tax to be paid, to leave ‘retained’ profit for the year, or drawings from retained profit reserves. If Mr Counts had decided to pay a dividend and was due to pay some tax, it would have looked like this, and in the closing balance sheet his retained profit reserve would have been reduced by £10,000.

Appropriation Account 31 March 2012

	Net profit (loss) after interest brought down	(5)
<i>less</i>	Tax due	3
<i>less</i>	Dividends declared	<u>2</u>
<i>equals</i>	Retained profit (reduction)	(10)

The Balance sheet

The balance sheet can usefully be shown with the opening and closing balance sheets for the year as the differences between them can be helpful for preparing the cash flow statement. The creditors can be combined with the debtors and stock to show net working capital. The balance sheet is also useful in terms of seeing the capital structure of a firm. Mr Counts has a debt to debt + equity ratio of 1/3, and with the retained profits as well, it is clear that Mr Counts is financed mostly by shareholders money.

The following balance sheets show that although Mr Counts made a loss of £5000, the cash position has moved from an overdraft of £10,000 to a surplus of £2500. This is why the cash flow statement is so important because the profit and loss account is based on revenues and expenses, not receipts and payments. Many revenue and expense transactions do not involve the movement of cash, such as depreciation charges. The cash movement occurred when the asset was first bought. A cash flow statement can be presented in terms of sources and uses of cash but it usually has a standard format in line with accounting standards. Both examples are shown below:

The Balance Sheets for year 3 (£000s)			
		BSo (2011)	BSc (2012)
Fixed assets			
Land		20	30
Other fixed assets – acquisition cost	100		100
<i>less</i> Depreciation provision	<u>40</u>		<u>60</u>
<i>equals</i> NBV		60	40
Working capital			
Stock – materials		20	16
Debtors	15		10
<i>less</i> Provision for doubtful debts	<u>0</u>		<u>2</u>
<i>equals</i> Net debtors		15	8
Prepayments		0	0.5
Trade creditors		(10)	(5)
Accrued expenses – wages		(0)	(2)
Cash at bank (overdrawn)		<u>(10)</u>	<u>2.5</u>
<i>equals</i> Capital employed		95	90
financed by			
Share capital	50		50
<i>plus</i> retained profits	<u>20</u>		<u>15</u>
<i>equals</i> Shareholders funds		70	65
Long term loans		<u>25</u>	<u>25</u>
<i>equals</i> Total finance		95	90

The Cash Flow Statement

The first statement is to reconcile the uses and sources of funds in the balance sheet. All of the items in the statement come from the differences between the opening and closing balance sheets, such as ‘change in stock holdings’, which is $20 - 16 = 4$. Given the cash flow statement has to adjust for transactions not involving the movement of cash, examples drawn from Mr Counts information for year 3 would be:

<u>Cash movements</u>		<u>Non-cash movements</u>	
Purchase of land	dr fixed asset a/c	depreciation	dr depreciation expense a/c
	cr cash a/c		cr depreciation provision a/c
wages paid	dr wages expense a/c	doubtful debts	dr doubtful debts expense a/c
	cr cash a/c		cr doubtful debts provision a/c

Sources and Uses Statement for year 3 (£000s)

Opening cash balance		(10)	opening and closing balance sheet calculations
Sources of cash			
<i>plus</i>	depreciation charges	20	60 - 40
<i>plus</i>	reduction in stock balance	4	20 - 16
<i>plus</i>	reduction in debtors	5	15 - 10
<i>plus</i>	increase in doubtful debts provision	2	0 - 2
<i>plus</i>	increase in accrued expenses	<u>2</u>	0 - 2
		33	
Uses of cash			
<i>minus</i>	retained (loss)	5	20 - 15
<i>minus</i>	capital expenditure	10	20 - 30
<i>minus</i>	reduction in creditors	5	10 - 5
<i>minus</i>	increase in pre-payments	<u>0.5</u>	0 - 0.5
		<u>20.5</u>	
<i>equals</i>	Closing cash balance	2.5	

The typical cash flow statement is usually presented as follows, and usefully brings in the operating profit for the year:

Cash Flow Statement for year 3 (£000s)

	Operating profit	(2.5)
<i>plus</i>	depreciation charge for the year	20
<i>plus</i>	Accruals adjustments	
	Creditors	
	Trade creditors	(5)
	Accrued expenses	<u>2</u>
		(3)
	Debtors	
	Trade debtors	5
	Provision for doubtful debts	2
	Increase in pre-payments	<u>(0.5)</u>
		6.5
	Stock reduction	<u>4</u>
<i>equals</i>	Total working capital adjustments	<u>7.5</u>
<i>equals</i>	Cash generated from operations	25
<i>less</i>	Interest paid	<u>2.5</u>
<i>equals</i>	Net cash available for investment	22.5
<i>less</i>	Acquisition of fixed asset	<u>10</u>
<i>equals</i>	Net cash increase (decrease) for year	12.5

Reconciliation

Opening cash balance	(10)
Net cash change for year	<u>12.5</u>
Closing cash balance	2.5

In terms of paying a dividend, it would depend on the 'dividend policy' of Mr Counts. He has sufficient cash and retained profits to be able to pay a dividend, but as he made a loss it might not be considered prudent to pay a dividend now.

Revaluation, historical cost (HC) and modified historical cost (MHC)

As accounts can be revalued to make fixed assets align with current replacement costs and realisable values, rather than the historical cost of their purchase price, a new statement has been introduced 'The Statement of Total Recognised Gains and Losses'. The following simple example illustrates this, with two years of historical cost accounts compared with a set of modified historical cost accounts for year 2. The asset cost 100 with a five year life, funded by equity capital, and depreciation is on a straight line basis. The modified historical cost accounts are for the asset being revalued by 100% at the end of year 1. A dividend of 10 is paid each year and sale prices are set to give return on capital employed (ROCE) of 20%.

	Historical cost a/cs		Modified historical cost a/cs	
Profit and loss a/c	1c	2c	1c	2c
Revenue	40	42	40	78
less Depreciation	<u>20</u>	<u>20</u>	<u>20</u>	<u>40</u>
equals Operating profit	20	22	20	38
less Dividend	<u>10</u>	<u>10</u>	<u>10</u>	<u>10</u>
equals Retained profit	10	12	10	28
Balance sheet	BS1	BS2	BS1	BS2
Fixed assets	100	100	200	200
less Provision for depreciation	<u>20</u>	<u>40</u>	<u>40</u>	<u>80</u>
equals Net book value (NBV)	80	60	160	120
plus Accumulated cash	<u>30</u>	<u>62</u>	<u>30</u>	<u>98</u>
equals Capital employed	110	122	190	218
ROCE (10% on NBVo + cash)	20%	20%	20%	20%
Financed by				
Share capital	100	100	100	100
plus Retained profits	10	22	10	38
plus Revaluation reserve	<u>0</u>	<u>0</u>	<u>80</u>	<u>80</u>
equals Shareholders funds	110	122	190	218

Statement of Total Recognised Gains and Losses

Movement in shareholders funds

	1c	2c	1c	2c
Operating profit	20	22	20	38
Revaluation	<u>0</u>	<u>0</u>	<u>80</u>	<u>0</u>
<i>equals</i> Total Recognised Gain	20	22	100	38
<i>plus</i> Contributions				
Equity capital	100	0	100	0
<i>less</i> Dividends paid	10	10	10	10
<i>equals</i> Movement in Shareholders Funds	+110	+12	+190	+28

The revaluation reserve is a 'capital' reserve from which dividends cannot be drawn until the reserve is 'realised' over time as dividends pass through the P/L account.

5. Economic profit, historical cost and revaluation, ratios

Economic profit

The profit and loss account has been shown to have a sequence of various types of profit classification, from gross profit through to retained profit. But does any of the accounted profits relate to a measure of economic profit, a measure which should ensure that it is based on including the full set of attributable costs? We know where to start, being the three costs of business: operating costs, capital consumption and the cost of capital finance. The problem arises with the cost of capital finance. Operating profit is before finance, net profit includes debt finance but not the cost of shareholder finance, and retained profit is after dividends paid, which, because they are 'discretionary', may be higher or lower than the shareholders long term cost of capital (r^*). The calculation of economic profit is therefore best calculated as operating profit less the net book value of capital employed times the cost of capital r^* , that is operating $\pi - \text{NBV} \cdot r^*$. The advantage of this calculation is that it shows the profit out of which additional dividends could properly be paid. Also it links well with economic analysis of present values, since the discounted sum of a series of economic profits equals the NPV of the associated cash flow series. The idea of economic profit links well with the accounting performance ratio, return on capital employed (ROCE) because $\text{ROCE} - r^*$ is the rate of economic profit or loss for the year.

Normal and supernormal profit

Normal profit is the position where there is sufficient revenue to cover the three costs of business – operating costs, capital consumption (depreciation) and the cost of capital finance (interest on loans and dividends to shareholders, the owners). This profit is sufficient to keep the company in business, and reflects the steady state position of a perfectly competitive economy, where all companies earn a normal profit. In practice, of course, the economy is 'dynamic', rather than in a perpetual steady state, and some companies will be earning more than a normal profit, and some will be making a loss. Since the three costs of business have to be covered before a 'real' profit can be earned, 'economic' profit reflects the profit over and above the normal profit, and an 'economic' loss when below it. Given normal profit is the benchmark for defining economic profit or loss, it should be of no surprise to find that at a normal profit, the present value of the associated cash flow series is zero. Economic profit would show a positive NPV and an economic loss a negative NPV.

Economic profit is often referred to a 'supernormal' profit, and in some cases as 'abnormal' or excess profit. To interpret supernormal profit it is necessary to rely on 'normative' economics. The economic profit could be 'good' profit which results from a particular firm's productivity and innovation. Alternatively, it could be 'bad' profit resulting from the abuse of monopoly power. However, monopoly is not necessarily a bad thing where that monopoly has been gained from productive efforts which customers find serves their interests, and where it continues to serve that interest. Where not, that monopoly power should be reduced in the public interest.

Historical cost and revaluation

The traditional approach to capital accounting in the profit and loss account and the balance sheet has been 'historical' cost. If you bought a fixed asset for 100, then its gross cost would remain that in the fixed asset account until it was fully depreciated and written out of the books. The published accounts do not, however, show a fair picture of the assets of the organisation if, over time, the relative cost or value of that asset has changed. In addition, the depreciation charges based on historical cost may not be a fair measure of the capital consumption value actually incurred by using the asset. Hence wrong decisions might be made.

Accounts based on current costs and market values may therefore be more informative, and the double entry system allows this to take place. Depreciation charges would then be based on the current value of the asset, not its historical cost. The first step is the revaluation of the historical cost of the asset. The double entry requirement means the formation of a new account in shareholders funds, being a capital reserve a/c. Suppose the current cost of such an asset had risen to 200, then the double entry would be:

Dr	Fixed asset account	100
Cr	revaluation reserve account	100

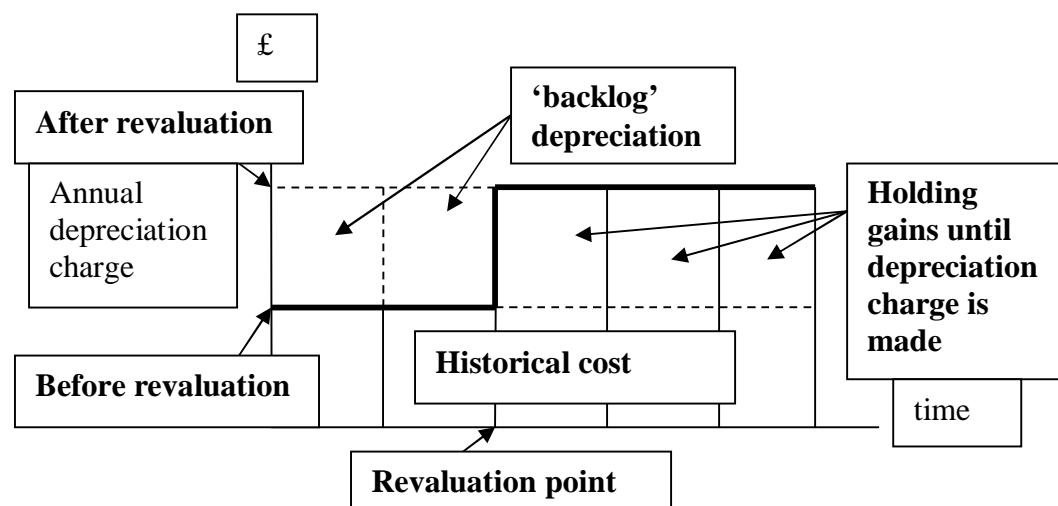
The balance sheet in its horizontal form would then show an increase of 100 on the asset side (dr), and equally 100 on the liabilities side (cr). A problem would then arise for the provision of depreciation, which must be adjusted to the new gross value. Suppose the asset was halfway through its life, then with straight line depreciation its historical cost provision for depreciation would be 50, giving a net book value of 50. Given a gross value now of 200, then the provision for depreciation should equal 100. The necessary increase in the provision for depreciation is provided from the newly formed capital reserve. The double entry would be:

Dr	revaluation reserve account	50
Cr	provision for depreciation a/c	50

and this means the provision for depreciation would now equal 100 and the net book value would equal 100 for an asset valued at a gross cost of 200. This technique in revaluation is known as accounting for the 'backlog' in depreciation. This can be presented in a chart below.

Revaluations in balance sheets require a new statement to accompany the traditional profit and loss account. This is because the revaluations create a 'holding gain' which is not passed through the profit and loss account, and in the chart above would be represented by the area shown. These are accounted for as 'unrealised' gains until each year's revalued depreciation charge to the profit and loss account converts them to 'realised' gains. The 'Statement of Total Gains and Losses' would incorporate the traditional profit and loss a/c figure and the holding gains or losses. This reflects the fact that the revaluation reserves created in shareholders funds increase their funds and therefore should be reported as gains (realised or unrealised) in the financial statements.

Revaluation gains and 'backlog' depreciation



Substance over form

Accountancy can therefore be quite sophisticated in its attempts to provide more meaningful information in its accounts, which are used for accountability and performance measurement. A move from historical cost to current cost accounting might be seen as an effort to show 'economic' substance over 'legal' form, and substance over form has become one of the guiding principles for setting accounting standards. Another example of this is the 'capitalisation of leases'. When you lease an asset you know that the provider is setting you a leasing charge which covers both the depreciation of the asset and its funding cost. Given the capital cost of the asset, the leasing charge would normally be the equivalent annual charge over the life of the asset. To make accounts comparable between organisations which lease and those which purchase the asset, those leasing are required to capitalise that lease in the balance sheet. The double entry system therefore requires a debit to capitalise the leased asset in the balance sheet and an equivalent credit to create the lease repayment liability account. These reduce over time as the expense charges are made to the profit and loss account.

Financial ratios

The accounts provide information which can be used to calculate ratios, and these can be measured against benchmarks in order to measure performance. This is important for investors in stocks and shares. A few of the key ratios arising from each of the main accounts are:

- * **Profit margin** – operating profit/turnover
- * **Interest cover** – operating profit/interest
- * **Price/earnings (P/E) ratio** – share price/earnings per share
- * **Earnings yield%** - earnings per share/share price
- * **Total shareholders return** (dividends pa + change in share price)/share price
- ROCE** – operating profit/capital employed
- Liquidity** – current assets/current liabilities
- Shareholders return** – earnings/net assets
- Gearing** – long term debt/equity plus long term debt
- Leverage** – long term debt/equity
- * **EBITDA** – earnings before interest, tax, depreciation and amortisation

Some of the ratios are important in financial analysis, concerning risk. If a company has a lot of long term debt compared to shareholders equity, it is highly geared or leveraged, which can be good for shareholders when business times are good, but this has a downside when times are bad because you are obliged to meet your interest payments and debt obligations, and if you cannot then the creditors can file for bankruptcy. With shareholder capital you are not obliged to pay dividends, and share capital is not repayable. In the long run you need to provide a fair return to shareholders, or you could not raise more money, but in the short run it allows the management of hard times. The prudent debt to equity ratio might be judged to be 40/60.

6. Management accounting

A distinction is made between financial and management accounting because one is concerned with the financial reports which are made available to those outside the management of the organisation (accountability), whether that be shareholders, banks, or other interested parties, whilst the other is the accounting information used in managing the organisation (control). Whether an organisation provides goods or services, it is clearly vital to know how much they cost, so that good management decisions can be made in terms of pricing and production. The skill is to know what one means by cost, and how best it might be calculated.

There are two major aspects to management accounting, first, 'product costing', and second, 'budgeting'. The important features of both is the 'behavioural' impact which can occur

because costs are properly classified to the goods and services in question, and the way in which budgeting can be used to provide useful incentives and control mechanisms. Management accounting inevitably becomes quite complicated because organisations are divided into many types of activity which support the goods and services which the organisation supplies, and for which it is particularly important to link the selling price to the cost of production. Managing in the short run might involve a different perspective on cost than managing for the long run. Budgeting is a well understood feature of everyone's lives, and its usefulness as a controlling mechanism is clear. Government's will typically have an annual budget review which looks at the on-going balance of public expenditure and taxation; a highly political process. Organisations will typically operate by giving their various divisions or departments' budgets, and have some mechanism for judging how they are performing with respect to their budgets.

Given the importance of management accounting as a decision control process, the importance of integration between economics and accounting becomes clear. There is a tendency in management accounting to ensure that costs are 'allocated' but whether the resulting information is economically useful is another question. Economists have often accused management accountants for being overly concerned with total and average costs, with little regard for marginal and opportunity costs, the important element required for making 'optimal' decisions. This need not be a problem however as long as managers know on what basis the accounting information is provided, and have the capability to make the necessary economic adjustments as required.

Product costing

The complexity of product costing is made clear by the various ways in which costs are classified. The first is to distinguish 'fixed' from 'variable' costs, whereby fixed costs are those which do not vary with output, whilst variable costs are those that do. However, this classification strictly only applies in the 'short run', given that fixed costs can change over time. The rent on a factory may be fixed for a year, so the cost of what you produce in the year is not affected by how many items you produce in the factory, but next year the rent could go up. In this way short run fixed costs can become 'long term' variable costs. Another important distinction is between direct and indirect costs, with indirect costs often referred to as 'overheads'. The difficulty of product costing is increased where more than one product is involved. In this case certain overhead costs, or 'common' costs, such as the cost of the legal department, might have to be 'allocated' to the various goods and services provided in an organisation. The allocation of 'joint costs', where an activity or process produces more than one output but in 'fixed' proportions, can be a quite complicated procedure, as can the allocation of service department costs, where service departments also provide services to each other.

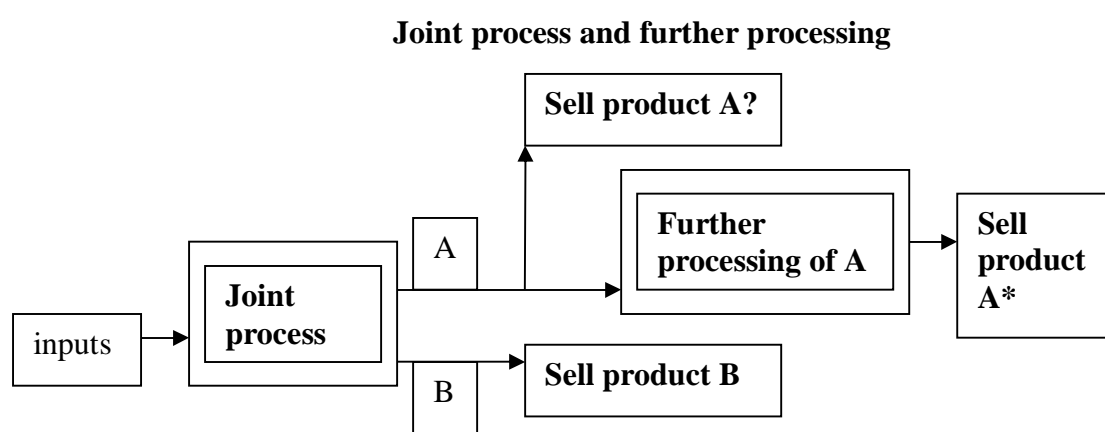
The first step in most textbook presentations of product costing is the 'break-even' point for an organisation in terms of sales output per year. Given the fixed costs for the year (FC), the price per unit (p), and the variable costs per unit (v), then the 'contribution' per unit will be $p - v$. Break-even will be where total revenue equals total costs ($TR = TC$), so $pQ = vQ + FC$, and by rearranging $Q = FC/(p - v)$. The break-even quantity (Q) is therefore given by dividing the fixed costs by the contribution per unit. The term contribution is a useful one because it can be referred to as contribution to 'profit' (which is important when considering what is the lowest price that you could set to stay in business in the short run) or contribution to meeting the fixed costs.

allocation and absorption

The allocation of fixed costs to particular products is referred to as 'absorption' of those costs into the products. A decision has to be made on which absorption 'base' to use, and in multi-product firms the choice of absorption base will affect the relative allocation of costs to

products. The calculations of the ‘recovery’ rate are usually made on the forecasts for the year, and can be based on either total output expressed as physical output (q) or financial value (£), total labour hours or total machine hours. So, for example, dividing the fixed costs by the total budgeted labour hours for the year will yield the ‘recovery rate per labour hour used’ (rL). So, if it takes ten labour hours to produce a good for sale, and rL had been accounted for as £7, then £70 of the fixed costs would be allocated to each unit of the good produced. It is usually the fixed costs which make it seem to the consumers that the cost of a product is higher than they would expect.

Joint costs in processing are another important feature of product costing. Where multiple products are produced from a joint process ‘in fixed proportions’ then the traditional allocation method based on a recovery rate is used, either using the total physical output of the joint process (eg, kilos) or the total realisable value of the products (£). Realisable value is the preferred method because realisable value is a much better indicator than weight for commercial decisions. The disadvantage of the physical output method is also that one or more of the products from the joint process might be accounted for as loss making (eg, a very heavy weight but low sale value, therefore the absorbed fixed costs are large and exceed the revenue from sale). The product costing of joint process outputs is also affected by whether there is ‘further processing’. In the following chart this situation is shown, and the question arises as to whether the allocation of joint process costs to products A and B should be influenced by the direct costs of further processing A to A* which only apply to product A.



The allocation of joint costs to products is often deemed by economists to be unnecessary because the important decision is whether to produce or not, and whether to further process or not. The decision to further process can simply be made by seeing whether the cost of further processing is exceeded by the additional incremental revenue from selling A* rather than A at the ‘split-off’ point. Once that decision is made, then it is simply a matter of comparing the cost of the joint process, plus the input costs to the revenue of selling A* and B. Nevertheless, it is often useful to have an ‘acceptable’ cost to a product, and certainly in accounting for the cost of stock and the calculation of profit or loss for the year.

The allocation of costs becomes more economically relevant when cost and price are brought together through demand. As products A and B will have separate demand curves, then the question arises as to whether there is an allocation of joint costs which would be profit maximising. This question is looked at further in the section on analytical examples – economic allocation of joint costs. In this an important distinction will be made between ‘primary’ products, which determine the overall production level, and ‘by-products’, which do

not. It is often thought that by-products should be sold cheaply, but the answer will be that they should be sold to the point where their marginal revenue is zero.

Activity-based costing (ABC)

Activity-based costing is well covered in most management accounting and financial control textbooks. The main purpose of ABC is to divide up the overheads and indirect costs into smaller activity sets, or 'cost pools', and to relate each of those cost pools to a particular causal factor (or 'cost driver'). If this is reliable, then the overhead costs allocated to different products can more accurately take account of the costs which result from producing each of the products, given that the different products use different amounts of the respective cost drivers. Traditional overhead cost allocations using labour or machine hours tend to be a gross average, and do not reflect the diversity of inputs required in a multi-product environment, or the costs arising from short production runs of specialist products. The idea of ABC is therefore to reduce any cross-subsidy in the cost allocation between products, thereby better reflecting the costs that should be attributed to each product. The calculation will therefore be a recovery rate for each cost driver used in producing the products and the costs absorbed by each product will be the sum of the amount of each driver used by a product times its respective recovery rate (eg, cost pool 1 divided by cost drivers 1 = recovery rate 1 times number of cost drivers 1 used by product A = allocated costs to product A from cost pool 1, and so on for cost drivers 2, 3...).

The 'economic order quantity' (EOQ) model is a good example of the use of ABC since it incorporates the cost of reordering and the cost of storing stock for sale. The overall purpose of EOQ is to minimise the overall cost of stock, taking account of these two costs, and given there is a trade-off between them (reordering frequently reduces the amount of stock held, thereby reducing the holding cost, but extra ordering increases the cost of that activity. If reordering is reliable then ordering 'just in time' is clearly efficient, compared with holding a large amount of stock, but JIT has a risk if it is not reliable, as stock may run out and customer demand not be met. It would be normal practice to hold a 'minimum safety stock'. The calculations involve four elements:

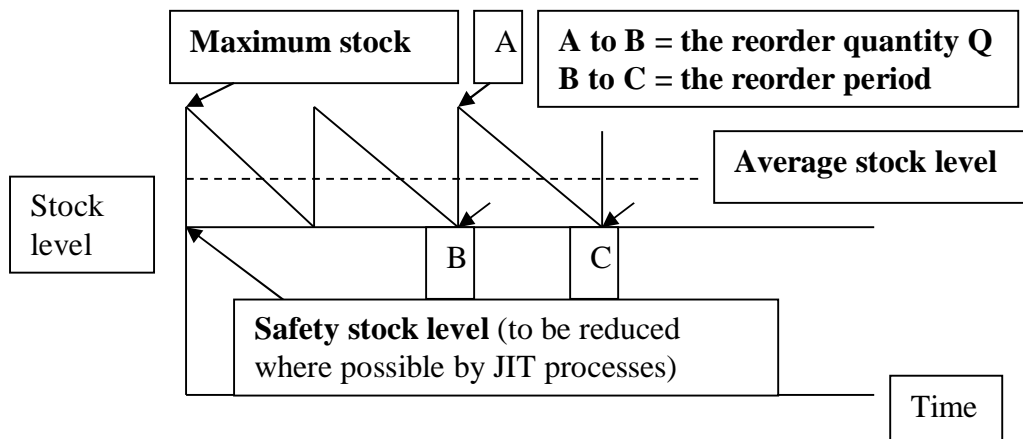
D = annual demand for stock; Q = the reorder quantity

C = annual cost of storage per unit of stock; R = cost per reorder made.

The total cost of storage is based on the average stock level (excluding safety stock, which is a long run fixed cost), that is, $(Q/2)C$, and the total cost of ordering is based on the annual demand for stock, that is $(D/Q)R$. The total cost of stock is therefore $TC = (Q/2)C + (D/Q)R$, and the total cost is a function of the reorder quantity (the higher the reorder quantity, the higher is the cost of storing the stock, but the lower is the cost of reordering because it is done less often).

The total cost is at a minimum with respect to the reorder quantity Q where the marginal cost (dTC/dQ) is set equal to zero, the point at which the positive marginal cost of reordering is cancelled out by negative marginal cost of storing stock, that is, the intersection point of the two cost curves for each activity. By differentiation, $C/2 - DR(1/Q^2) = 0$, therefore $Q^2 = 2DR/C$ and so the optimum reorder quantity is $Q = \sqrt{2DR/C}$, in other words, the square root of two times the annual demand multiplied by cost per reorder made divided by the annual cost of storing a unit of stock. The outcome can be represented in a chart:

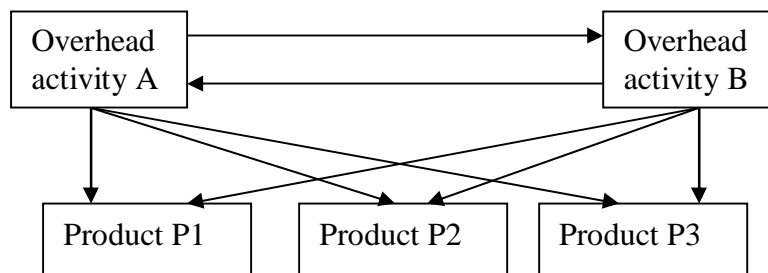
The 'economic order' quantity model



Overhead allocation

The traditional methods of overhead allocation may therefore be basic, such as simply having a standard % uplift on the cost of each good or service produced to recover the cost of the overhead services, such as power supply, maintenance, personnel departments and so on. They can be more 'accurate' by identifying what activities make particular overheads important to have, and identifying the 'cost drivers' which result in the level of overhead costs which emerge. Each good or service produced then bears more of its fair share of the overhead costs it causes. Where service departments (say A and B) also serve each other, as shown in the chart, the allocation of overheads to the goods and services produced (say P1, P2 and P3), requires some attention, unless of course, it is simply ignored and A and B overheads are simply allocated across P1-3. Two methods are generally employed, first the 'step-down' method, and second, the 'reciprocal' method.

Allocating overheads to products



The step-down method can be illustrated for three service departments (S) and two products (P) as follows, giving first the percentage of its costs, out of 100%, which each service department provides to the producing departments and other service departments. The step-down process then starts by selecting one service department first and allocating its costs, and so on, each service department not receiving any more costs once it has allocated its costs. At each step the percentages have to be adjusted (eg, m is then shown as m*) because the service departments dealt with in each of the previous steps cannot have costs allocated to them:

Percentage of service supplied by each service department

	P1	P2	S1	S2	S3	
S1	a	b		c	d	= 100%
S2	m	n	q		r	= 100%
S3	s	t	v	w		= 100%

Accumulation of allocated overhead costs to producing departments

	P1	P2	S1	S2	S3
Step 1 Costs of S1= C1	C1a	C1b	(C1)	C1c	C1d
Step 2 Costs of S2= C2+C1c	C2m*	C2n*		(C2+C1c)	C2r*
Step 3 Costs of S3= C3+C1d+ C2r*	<u>C3s*</u>	<u>C3t*</u>			(C3+C1d+C2r*)
Total costs allocated at end	TP1	TP2	0	0	0

This method is clearly making an attempt to be more accurate than the straightforward 'direct' method, but still may not reflect the best possible allocation which fully accounts for the services supplied between service departments. This relies on using the simultaneous equation approach of linear algebra. Taking two service departments with costs SC1 and SC2 (and total costs after allocation of M and P respectively) supplying each other and two producing departments (A and B), we could set out the total costs of the two service departments, given the percentage of service supplied to each other (0.05 and 0.2) as:

$$M = SC1 + 0.05P$$

$$P = SC2 + 0.2M$$

By substitution of P into M we get $M = SC1 + 0.05(SC2 + 0.2M)$. Given from this we can calculate M, then $P = SC2 + 0.2M$. Having calculated the total costs for P and M, they can be allocated to the producing departments A and B.

The last aspect of cost allocation to consider here is where an organisation has separate divisions, each of which has the choice to choose an outside supplier of a service (say cleaning or catering), or to use the 'in-house' supplier. In situations where the cost of supply is lowest when all divisions choose the in-house supplier, then management might best decide to allocate the in-house overhead in such a way that it is not in the interest of any division to choose the outside supplier. Such a system was suggested by Moriarity to motivate divisional managers to use internally provided services, and in effect meant that the overhead allocated to each division would be less than the tendered cost for an outside supplier.

7. Budgeting

The purpose of budgeting is to provide a connection between the future and the past, and to create a framework within which management can make decisions and control the process. Budgets are forecasts as well as benchmarks. We may budget to spend £x, but find we spend more, or the budget of £x might be fixed in such a way that more cannot be spent without getting permission to do so. It therefore plays a key role in effective management, and the accounting part of that is 'variance' analysis, whereby the budget forecasts are compared with the figures that actually result, the 'actuals' or outcomes.

Variance analysis has developed into a reasonably complex set of standard calculations, and this will be explored further in one of the sections on the applications of political economy. However it is worth illustrating variance analysis at this point to show that standard computational rules are required to ensure that arithmetical consistency is maintained. This is because different variances need to be added together to make other variances, all possible variances being brought together in a 'decomposition' chart. Let us take the example of the 'total revenue' variance and its split into the 'price' and 'quantity' variance. The adding up rule is that:

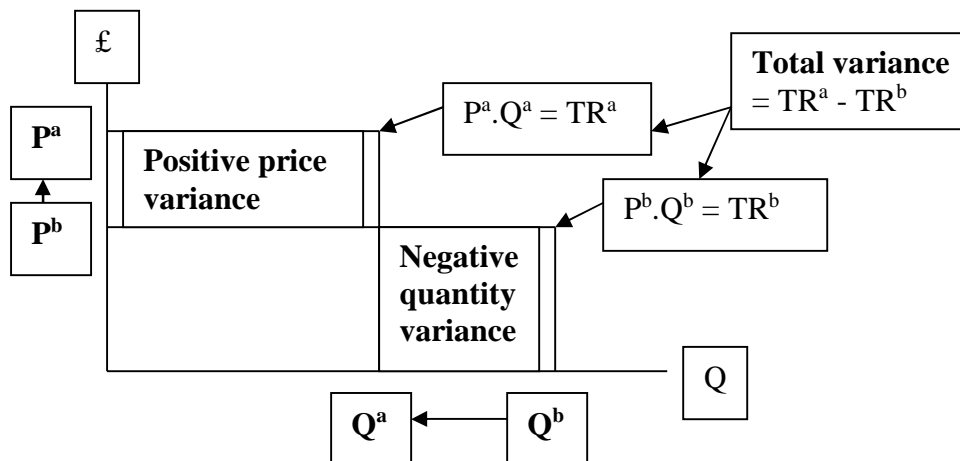
- * **Price variances are based on ‘actual’ quantities**
- * **Quantity variances are based on ‘budgeted prices**

The chart shows that a budgeted total revenue at the start of the accounting period ($P^b.Q^b$) becomes an actual total revenue ($P^a.Q^a$) at the end of the period. In this case it shows that the price was increased compared to the budget price, but the quantity sold fell, compared with what had been budgeted to sell. The two variances are therefore:

- * Price variation = $(P^a - P^b)Q^a$
- * Quantity variation = $(Q^a - Q^b)P^b$

It is helpful in the calculations to put the variation (ie, change Δ) you are interested in first, then multiplied by (ie, evaluated at) the actual quantities or budgeted prices. In this particular case the quantity variance is negative, or ‘adverse’, and the price variation is positive, or ‘favourable’. The two added together will equal the total revenue variance $TR^a - TR^b$.

Comparing Price and Quantity Variances



It is useful to mention that variance analysis, as is budgeting, is a key component of the practice of ‘responsibility’ accounting. This links the use of variance accounting numbers with performance measurement of individuals, divisions or organisations, and the technique is not necessarily straightforward in application. To take a simple example, a manufacturer might have divided responsibility for purchasing of materials from the processing of materials in a workshop, with a separate manager responsible for each. Each might have been given a budget for their part of the production process, and the first step in variance analysis is to look at the difference between the budget and the actual outcome for each, and to identify whether it is a ‘positive’ or ‘negative’ variation. Suppose the purchasing manager had a positive variation, but the workshop had a negative variation. Should we necessarily conclude that the purchasing manager has done a good job – and deserves a bonus – and that the workshop needs to get itself in order. Not necessarily, because the variances might be causally related, and one manager is reliant on the performance of the other. In our case, suppose the positive purchasing variance had arisen because the purchasing manager had opted for cheaper materials, and imposed lower quality materials on the production manager. The negative production variance might then arise simply because the workshop had to use more of the lower quality materials to produce the goods necessary. In effect, both variances should be taken and judged together if management is going to control an organisation fairly and effectively. The answer to such things is that

variance analysis is a process of discovery, for management to ask why the variances occur, how that might best be avoided in the future, and to learn from the past for the benefit of better actions in the future.

The budgeting process, and the use of variance analysis, can therefore be a stressful process, and it is often convenient to argue for stability and the maintenance of existing budgets. To avoid such perpetual 'incremental' budgeting, the practice of 'Zero-based' budgeting has developed. The purpose is to go right back and ask why are we doing this, is it necessary, and can a better approach be found? This can be a threatening process where people or divisions might lose their job, or be changed to something in an unexpected way. The answer to this is sensitive management, to enable all to be aware of, and understand why, change is necessary, and for management to deal with the outcomes in a fair and considered way.

8. Corporate finance and the cost of capital

Throughout the course text, to illustrate the technique of discounting and compounding, and to calculate present and future values, the discount rate has been represented by r^* , the 'test' discount rate or 'cost of capital'. The cost of capital is the amount of return required by investors for them to finance a company's assets, whether by lending to the company or purchasing shares. Corporate finance is the specialist subject concerned with what r^* should be in practice, and how it might vary in different circumstances, but one key principle is that the required return has to reflect the 'risk' of the investment. The more risk there is, the more return will be required.

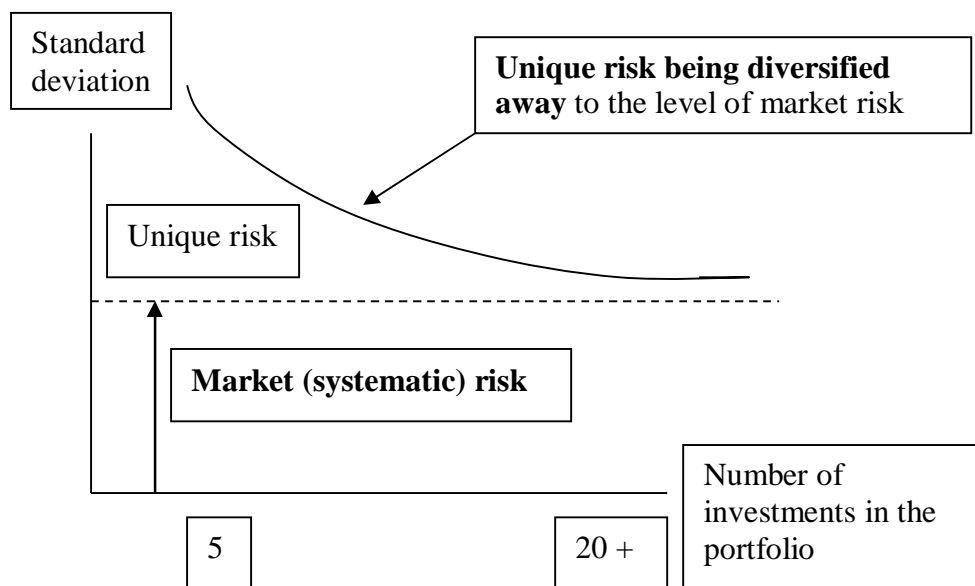
To put corporate finance in context, it is useful to consider the balance sheet. On the one side, the assets, and on the other side the finance liabilities. The normal presentation of published balance sheets being assets first (capital employed), which are 'financed by' shareholders funds and long term loans. The market value of the company's assets will depend on the expected future cash flows discounted at the opportunity cost of capital. This market value is expected to reflect all available information and respond quickly according to the 'efficient capital markets' theory. The consequence of this is that in well functioning capital markets, all equivalent risk assets are priced to offer the same expected return. Underlying this is the 'the law of conservation of value', which simply says that the total of two assets' cash flows is just the sum of the two assets (ie, $NPV\ of\ A + NPV\ of\ B = NPV\ of\ A + B$). This means that nothing is gained from a merger of two companies if its purpose is simply to add cash flows.

The analysis of risk depends on circumstances. An investment appraisal by a company will require an analysis of the risk of the new prospective asset's cash flow compared to the risk of all the company's existing assets combined. For investors in the company they will be concerned with the return offered on all the company's assets compared with the return offered from all other possible combinations of investments in other companies. This leads to measuring 'portfolio' risk because risk can be divided into two parts: the 'unique' risks applying to each company, which can be 'diversified' away by investing in more than one company, and market or 'systematic' risk which cannot be diversified away. A simple example would be deciding whether to invest in umbrella makers or ice cream producers. The unique risk of the weather to each company could be diversified away, but the market risk of an economic downturn affecting both could not.

Investing in just one firm would mean that you bear both the market and the unique risk, and therefore the required return would be high. By investing in both firms, the offsetting (correlated) variances for the companies relating to unique risk would reduce the overall unique

risk and therefore the required return would fall. The relationship between unique and market risk as the number of investments in the portfolio increases can be illustrated in a chart, where the standard deviation of expected market return (E_{r_m}) to actual market return (A_{r_m}) is plotted against the number of investments in the portfolio. Variance and standard deviations are the natural indices of risk in practice, the standard deviation being $1/n-1 \times \sum (A_{r_m} - E_{r_m})^2$. As the number of different shares in the portfolio increases, the return begins to tend towards the market risk, so the risk of a well diversified portfolio depends on the market risk of the securities included in the portfolio.

Unique and market risk compared

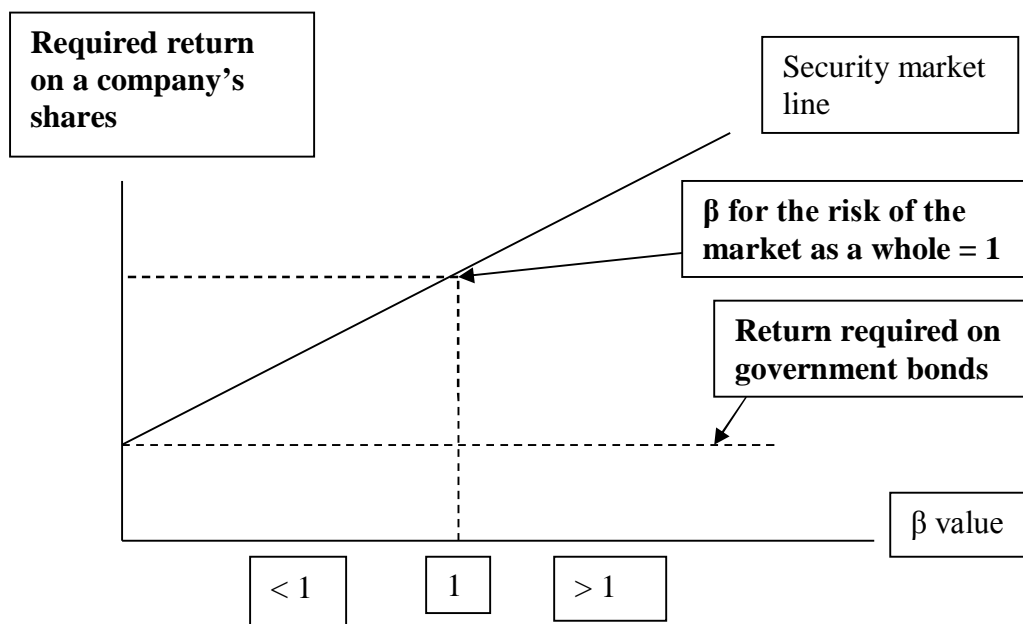


Corporate finance exists within a sophisticated market context, and the dynamics of such a market have to be taken into account. Given 'option' demand (ie, the option to buy or sell a share at a fixed price in the future), and the methods of valuing such options developed by Black and Scholes, the experience of the global financial crash which occurred in 2008 has shown how sophistication can lead to a very unstable outcome. The idea of 'efficient' markets has also to take account of human behaviour which tends to focus on recent events rather than long run trends, leading to a 'herd mentality' or 'cascade' effect.

Capital asset pricing model (CAPM)

The risk of a particular firm can now be compared to the market risk of a well diversified portfolio. This is the role of the capital asset pricing model which links the return required on the most riskless investment (government bonds) to the return required for the market as a whole and the return required for a company's particular portfolio of assets. The 'security market' line plots the expected return on an investment in a company against its 'Beta' (β), which is greater than one if the portfolio of a company's assets are more risky than the market as a whole, and less than 1 if less risky. The security market line therefore shows how the value of the firm is affected by a change in the aggregate value of all the assets in the economy.

CAPM and beta



Given a beta for a company, the expected risk premium on its shares is therefore: $\text{beta} \times (\text{market risk} - \text{government bond risk})$. Overall this means that there are four main values of r^* which can be calculated:

- * The 'risk free' rate based on government bonds
- * The market rate based on a balanced portfolio of shares
- * The company's cost of capital based on its portfolio of assets (ie, β)
- * The opportunity cost of capital for investment appraisal where the project's risk differs from the risk of the company's portfolio of assets

In the last case, this is often referred to as a 'hurdle' rate when discounting a project's expected cash flows. One lesson for a company is that because investors can eliminate unique risk by diversifying their asset portfolio, companies do not need to do that. One lesson for investors is that because they can lend or borrow at the 'risk free rate', then they should look for the portfolio of shares which offers the highest ratio of risk premium to standard deviation.

Calculating betas in practice is not simple, and there are other models which have been put forward as an alternative theory of risk and return. One is the Arbitrage pricing theory (APT), where the risk premium is based on 'factors' or 'macroeconomic' effects. In this case the formula would be a weighted sum of all the factors' effects. Another is the Dividend growth model (DGM). It links the dividend yield of a company plus the expected growth in dividends to the cost of equity, and is shown algebraically as $r_e = D/P + g$. In practice, whichever model is used (CAPM, APT or DGM), the key elements are that investors expect a higher return for taking risks, and are only concerned with market risks, not diversifiable risks.

Capital structure

The balance of shares to debt for a company is much discussed, and arises because with debt the company has to pay interest and repay the principal sum borrowed, and the debtors have first call on the company's funds, whereas companies do not have to pay dividends to shareholders or

repay their share capital. However, shareholders receive all the profits of a company after tax, which can either be paid as dividends or retained in shareholders funds.

A fully share financed company therefore risks losing the whole value of the company but if a proportion is funded by debt then the total risk is less to them and, if company performance is increasing, the increase in returns to shareholders is 'leveraged'. However, a high proportion of debt puts the shareholders at risk of bankruptcy if there is an economic downturn. It is often argued that the overall cost of capital to a company is minimised with a debt to equity + debt ratio of about 40%.

However, it has been argued, particularly by Modigliani and Miller, that because the value of a firm is related to its assets and expected cash flows, the division in financing between debt and equity should not change the total value of a company (the law of conservation of value). Alternative views exist because tax rates can differ between interest and dividends (interest receipts are often taxed but dividends are not because the company's profit after interest is taxed). Also, in behavioural terms, high levels of debt might incentivise company management to improve performance because of the fear of 'financial distress'.

In a similar way to debates on debt/equity ratios, there is a debate on whether dividend payments make any difference to the value of a company. In essence the argument that it makes no difference is based on the fact that as investors do not need dividends to have access to cash (they can sell shares), then they are unlikely to pay higher share prices for companies that have a high dividend payout. The dividend policy of companies depends on whether it is a 'mature' company, where a high proportion of earnings are typically paid out, or a 'growth' company where little or no dividends are paid. In general, dividends are only changed when there is a shift in a company's long term sustainable earnings and, because of the need to maintain market confidence, companies are reluctant to make dividend changes that might need to be reversed.

The policy debate on the cost of capital in UK regulated industries

Estimates of the cost of capital depend on the model being used and the outcomes of the current data used to predict it, given returns on company stocks and market returns vary, and risks themselves change over time. The debate on the cost of capital was well illustrated in the utilities and network industries privatised in the UK from 1984 onwards because it an important part of the price setting process by the 'economic regulators'. A report was commissioned by the UK economic regulators and the Office of Fair Trading in 2002 (A Study into Certain Aspects of the Cost of Capital for Regulated Industries in the UK by Wright, Mason and Miles) which noted that:

"The cost of capital is a key input into the regulatory process for utilities. In setting price limits regulators need to decide what would constitute a 'fair' rate of profit and therefore need to assess the return investors in these firms would have earned if they invested in any other firm with a comparable level of risk".

The purpose of the report was to "achieve greater consistency between regulators than rely on the ad hoc approaches used to date".

Different approaches are well illustrated in 1991, two years after privatisation, when the Director General of Water Services published a consultation paper on the cost of capital in the water industry, in preparation for the first five year periodic review in 1994. The response of the regulated water companies makes interesting reading (The Cost of Capital in the Water Industry – A response by the Water Services Association and the Water Companies Association, November 1991). The accompanying 'statement' said "The Director General has stated that the real cost of capital in the water industry is 5-6% after tax. The industry considers this assessment is too low and that procedures applied to arrive at this result are seriously flawed

both in terms of the principles involved and the way the evidence has been interpreted". It goes on to say that "A level of at least 9.5% is justified given the risks to which water companies are exposed", stating such things as:

- * Heavy programmes of investment to meet service obligations to customers
- * Unique risks such as regulatory and political intervention
- * Environmental impact rules.

and from their concern proposed a 'Joint Forum' with Ofwat to ensure that they can arrive at a reasonable solution based on consensus.

One reason for the difference of opinion was the different 'models' thought appropriate. The response suggests that Ofwat preferred the 'Dividend Growth Model' whereas the industry preferred the Capital Asset Pricing Model, stating that "Although Ofwat believes the CAPM approach produces 'somewhat implausible numbers', the industry's view is that it currently provides the best theoretical and practical approach to dealing with the cost of capital", and going on to say that "The UK incentive-based price-cap system of regulation inevitably involves risks for investors. The CAPM, which is concerned with investor choice under conditions of risk, is a much more appropriate model to use in these circumstances".

The 9.5% proposed was derived from a risk free rate of 4% (post tax) and a market risk premium on equities of 9% (the arithmetic mean of past annual returns on the market as a whole) with a water industry beta of 0.6 (therefore $9\% \times 0.6 = 5.5\%$). The beta reflected the sensitivity of returns from water and sewerage company shares to movements in the equity market as a whole, using data on share price movements since flotation in 1989.

In 1995 British Gas TransCo published a response to Ofgas on their consultative document on TransCo's price control. They noted that in the 1993 inquiry the Monopolies and Mergers Commission (MMC) had "estimated the range of TransCo's cost of capital was 6.5-7.5% pre tax". Their opinion was that the cost of capital was now in the range 7.4-8.9%, and that the specific reason for this change was the riskiness of TransCo relative to the stock market as a whole, their beta having risen to 0.83-0.93. It also reported the costs of capital established in other sectors, being water 6-8%, electricity distribution 7%, electricity transmission 6%, Railtrack 7.5%, BAA 7.5-8% and BT 15%, noting that the rate of return set by HM Treasury for nationalised industries was 8%.

The response made some interesting comments about the risk of being a regulated company. First, "All cost of capital estimates depend on the underlying assumptions and rely on past data. The risks to investors in TransCo are to a large extent determined by the by the form and content of the regulatory formula and the exercise of regulatory powers in relation to that formula". Secondly, in particular it stated that "more generally the CAPM cannot take account of the asymmetric nature of regulation and its effect on required returns...and regulation cannot compensate for investments becoming uneconomic and yielding a return below their cost of capital". This is because "returns are 'capped' in an upward direction....and means that there is a serious risk of underinvestment where there is a significant exposure to commercial and technological risk".

By 2006, the debate had moved forward and Ofwat and Ofgem published a discussion paper, Financing Networks. This arose because by 2004 there was concern about the number of regulated water companies with highly geared capital structures, and when Welsh Water was put into administration it was replaced by a company with no share capital, Glas Cymru. The concern was that regulators would be less able to protect consumers if highly geared companies

were to become subject to financial distress, in other words, risk would move from shareholders to consumers. However, both Ofwat's and Ofgem's periodic price controls were considered successful. The three main reasons were that:

- * Regulatory risk had been diminished, particularly because the regulators had become more transparent in their operations, and had put a lot of emphasis on 'City briefings' and communicating with market investors.
- * Both debt and equity markets had shown their willingness to finance the heavy investment programme
- * Efficiency had improved in the regulated companies and so price increases had been constrained.

One suggestion (by Helm and Meyer) to keep the cost of capital at the lowest level was to allocate the risk of capital to different funding sources, that is, have a 'split cost of capital'. Once an investment had been formally added to the regulatory asset value (RAV) by the economic regulator, it in effect becomes relatively low risk, and is suitable for the low cost of debt finance. Where investments are being planned and built, then the incentive properties of equity funding make it suitable for managing operating and construction risks. However, two problems which had arisen because of the timing mismatch between a regulatory five year price setting cycle and the longer term financing requirements of capital intensive infrastructure businesses were the 'financeability' and the 'regulatory commitment' problems. Approaches to tackling these problems could include increasing the price setting period from 5 to 10 years, or to set the cost of capital for a period longer than the five-year norm.

The documents published on the cost of capital for the regulated industries reflect the fact that in practice the cost of capital is a complicated, technical subject. However, the consultation process which is such a part of the regulation of the privatised utility and network industries has meant that there is a broader understanding of the issues that need to be addressed. In 1993 Ofwat published its 'framework and approach to the 1994 periodic review', and in the cost of capital section states that "the Director has reviewed the responses to its 1991 consultation paper", but "the Director is minded to maintain the broad thrust of the earlier consultation document in reaching final decisions on the cost of capital". In a similar publication for the 1999 periodic review it is noted that "The 1994 Periodic Review used a cost of capital of between 5-6% after business taxes" and that the Director "remains of the view that there is no single satisfactory approach to determining the cost of capital and will continue to assess evidence using both the Capital Asset Pricing Model and the Dividend Growth Model. In terms of continuity, "the Director might expect that the weighted average cost of capital would be unlikely to deviate substantially from that used in 1994". Again, in a similar paper for the periodic review 2004, Ofwat said that "it intends to follow the same approach as at the 1999 review for the capital base, and the regulatory capital value", and considers that "the water industry remains low risk in absolute terms and relative to other industries".

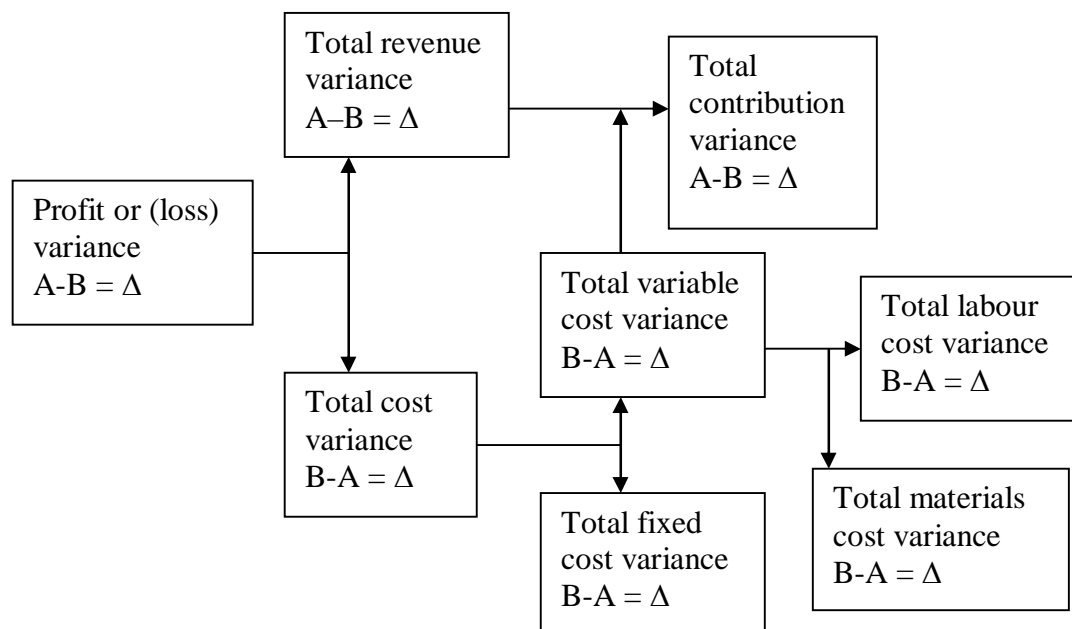
Applied Accountancy - Management Control and Performance Measurement

1. Variance Analysis

The analysis of variations between planned or expected outcomes, and the actual outcomes, is an important, practical management tool for economists and accountants, giving a focus on understanding the past and improving decisions for the future. Regulating utilities by the RPI – x system was a classic example, where the ‘budget’ is set by the regulator and the regulated company seeks to outperform that in relation to the outturn, and the variance analysis shows whether an economic profit was achieved. In practice, the general approach is to be found in management accounting text books. Three types of variance need to be taken into account: price variances, quantity variances and accounting variances which arise where fixed costs are treated as variable costs (absorption costing). The variances can relate to inputs and outputs, such as selling prices and sales quantities, or the price paid per unit of input into the production process, and the quantity of input per unit of output.

The basic model of variance analysis brings together economics and accounting, and focuses on profit or loss (π), total revenue (TR), total cost (TC), variable costs (VC), fixed costs (FC) and contribution (revenue minus variable costs). Variance is the difference (Δ) between forecasts for budgets (B) and the outturns (A). To keep the arithmetic right for variances, a + variance being ‘favourable’, and a – variance being ‘adverse’, it means that revenue variances are Actual minus Budget and cost variances are Budget minus Actual. The basic model could be shown as follows, with the arrows showing the related variances, and how they are split or combined, moving from left to right:

The main financial variances



The adding-up rule

From this it can be seen that there is a hierarchy of variances which results from ‘decomposing’ particular variances into a number of other variances. The consequence is that there has to be an ‘adding-up’ rule to ensure that two variances can be added in total to another variance. For example, the total revenue variance can be decomposed into the selling price variance and the sales quantity variance because total revenue is price multiplied by quantity. This can be used

to find the adding-up rule as follows, with superscripts B and A being budget (or forecast) and actual (or outturn) respectively:

$$\begin{aligned} \text{TR}^A - \text{TR}^B &= \text{TRV}^\Delta = \text{total revenue variance} \\ \text{and } P^A.Q^A - P^B.Q^B &= \text{TRV}^\Delta \end{aligned}$$

so to calculate the price and quantity variances we need to decide which multiplier each variation should be multiplied by. If price variations are calculated (evaluated) at 'actual' quantities and quantity variations are calculated at 'budgeted' prices, then added together they equal the above, that is:

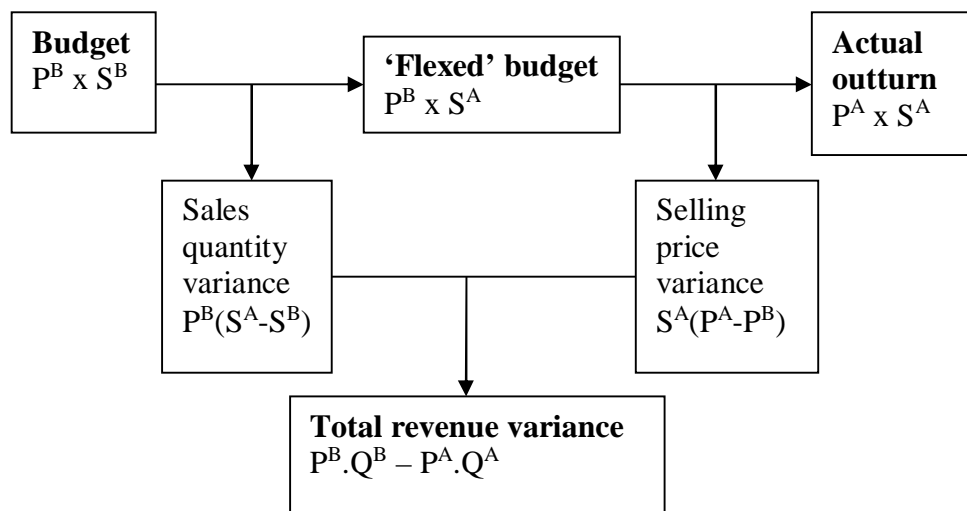
$$\begin{aligned} \text{selling price variance} + \text{sales quantity variance} &= \text{total revenue variance} \\ (P^A - P^B).Q^A + (Q^A - Q^B).P^B &= \text{TRV}^\Delta \end{aligned}$$

and an illustrative worked example shows that a variance can be 'favourable' (positive) or 'adverse' (negative), given the worked example is based on an actual price increase causing a fall in the number of sales:

$$\begin{aligned} \text{Selling price variance} &= (15 - 10)(20) &= +100 \text{ favourable} \\ + \text{Sales quantity variance} &= (20 - 30)(10) &= -100 \text{ adverse} \\ = \text{Total revenue variance} &= (15 \times 20) - (10 \times 30) &= 0 \end{aligned}$$

Accounting for the 'flexed' budget

The above variances have all been calculated as the difference between the budget figures and the actual figures. Management accountants have introduced a 'hybrid' position which enables more variances to be calculated, calculating a 'flexed' budget, which involves both budget and actual figures. The difference between the budget figures and the outturn figures is therefore 'partitioned' by an actual outturn figure multiplied by a budgeted figure. The following illustrates this based on the revenue variances above.



The cost side is introduced when budgeted costs are used to obtain the 'standard' cost of production per unit, and these are retained when taking account of actual sales compared to budgeted sales. In short, budgeted costs x budgeted sales = budgeted cost, and budgeted cost x actual sales = flexed budgeted cost. The next step is actual costs x actual sales. The introduction of the flexed budget enables:

* the **total revenue variance** to be split into the **selling price variance** and the **sales revenue volume variance**

* the **total variable cost variance** to be split into the '**standard**' **variable cost volume variance** and the **input price and efficiency variances**

* and therefore the **standard contribution variance** is the **sales revenue volume variance** plus the **standard variable cost volume variance**

The main formulae for this extension of the basic model through the 'flexed' budget can then be shown as follows, given first the notation:

P = price per unit of output sold

S = quantity of sales

I = inputs per unit of output

C = cost per unit of input

FC = fixed costs

Superscripts

A = actual

B = budget

Subscripts

L = labour

M = materials

Output variances

(flexed budget to actual outturn)

***Selling price variance** = $(P^A - P^B)(S^A)$

(budget to flexed budget)

***Standard contribution variance** = $(S^A - S^B)(\text{Std contribution})$

Input variances

(budget to flexed budget)

***Standard variable cost variance** = $(S^B - S^A)(\text{Std variable cost})$

(flexed budget to actual outturn)

***Materials usage variance** = $(I_M^B - I_M^A)(C_M^B)(S^A)$

***Materials price variance** = $(C_M^B - C_M^A)(I_M^A)(S^A)$

***Labour efficiency variance** = $(I_L^B - I_L^A)(C_L^B)(S^A)$

***Labour rate variance** = $(C_L^B - C_L^A)(I_L^A)(S^A)$

(budget to actual outturn)

***Fixed cost expenditure variance** = $(FC^B - FC^A)$

Absorption costing – an accounting model of variances

The need to take account of fixed costs when defining a standard profit, rather than a standard contribution, means that fixed costs have to be 'absorbed' in some way into the cost per unit of production. This can be achieved by choosing an 'absorption base' (say budgeted sales, or labour hours etc) and dividing the total budgeted fixed costs by the total budgeted absorption base. This yields the fixed cost 'recovery' rate (r) per unit of the absorption base used. If labour hours were chosen as the absorption base, and each product produced uses 8 labour hours, costing £5 per hour, then an example could be:

Total budgeted fixed costs = 400

Total budgeted labour hours = 800

Recovery rate = $400 \div 800 = 0.5 \text{ per labour hour}$

Standard cost per unit produced equals variable cost + absorbed fixed costs

therefore equals $8 \times (5 + 0.5) = £44$ per unit produced. If the budgeted sales are achieved at budgeted costs then the fixed costs will be recovered, given the units are sold for £44.

The problem with absorption costing is that it treats fixed costs as though they are variable costs, and if actual sales are greater than budgeted sales, then the fixed costs are 'over-recovered', and 'under-recovered' if actual sales are less than those budgeted for. The result is that a further variance has to be added in to account for this over or under recovery. This is the fixed cost sales volume variance and the relationship with contribution is, from the 'adding up' rule:

$$\begin{aligned} & \textbf{Standard contribution variance} \\ &= \textbf{Standard profit variance} + \textbf{Fixed cost sales volume variance} \end{aligned}$$

Final variances

The remaining variances in a 'full decomposition' set arise because 'mix' can be taken into account where there is more than one substitutable product produced or input available. Also, sales may not equal production, resulting in additions or subtractions from stock. Examples of mix variances which can be calculated where there are two or more substitute products are selling basic cars and deluxe cars, or using substitute inputs in production, such as wood or plastic. Account can also be taken of competitors, looking at market size and share. The formulae arising from absorption costing and these additional variances, given the additional notation, are:

Q = quantity produced

r = absorption rate per unit of input/output chosen as recovery base

ASP = actual at standard proportions r_L, r_M = input used as absorption base

EMS = expected share at actual market size

and the absorption base is labour hours (L)

Variances from absorbing fixed costs and where sales \neq quantity produced

* Standard contribution variance

$$\begin{aligned} &= \textbf{Standard profit variance} (S^A - S^B)(\text{standard profit per unit}) + \textbf{Fixed cost sales} \\ &\textbf{volume variance} (S^A - S^B)(r_L^B \cdot I_L^B) \end{aligned}$$

* Fixed cost production volume variance = $(Q^A - Q^B)(r_L^B \cdot I_L^B)$

$$\begin{aligned} &= \textbf{Fixed cost sales volume variance} + \textbf{Fixed cost stock production volume variance} \\ &\text{As above} \quad + \{(Q^A - S^A) - (Q^B - S^B)\}(r_L^B \cdot I_L^B) \end{aligned}$$

* and the Fixed cost stock production volume variance

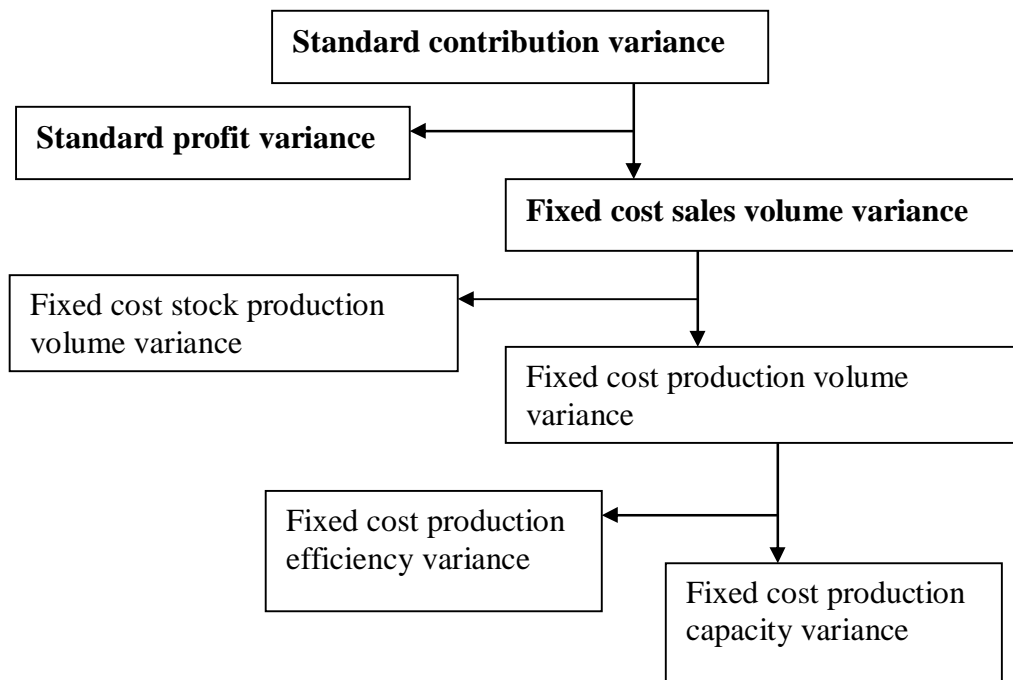
= **fixed cost production efficiency variance**

+ **fixed cost production capacity variance**

$$\text{which is} \quad (I_L^B - I_L^A)(r_L^B)(Q^A) + (Q^A \cdot I_L^A - Q^B \cdot I_L^B)(r_L^B)$$

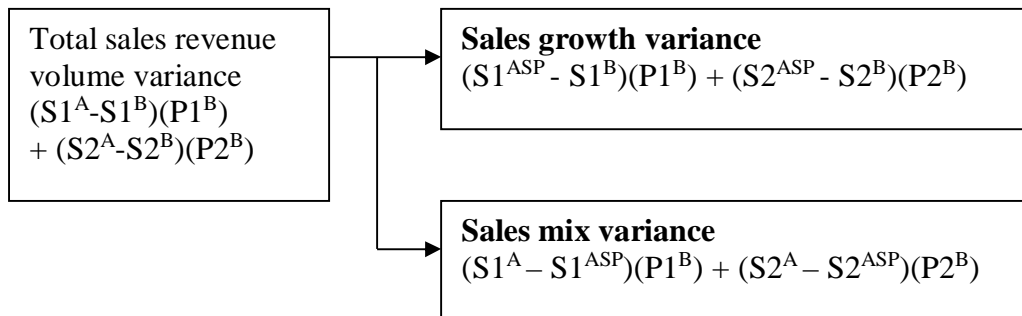
which all together can be shown, from top to bottom, as:

Absorption variances plus variances where $S \neq Q$

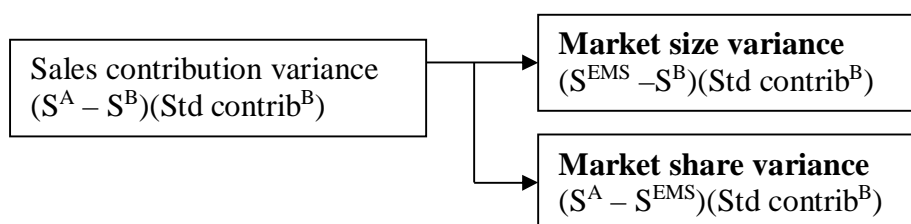


and the mix variance formulae are shown in the following diagrams, given two substitutable products or inputs, noted as 1 or 2:

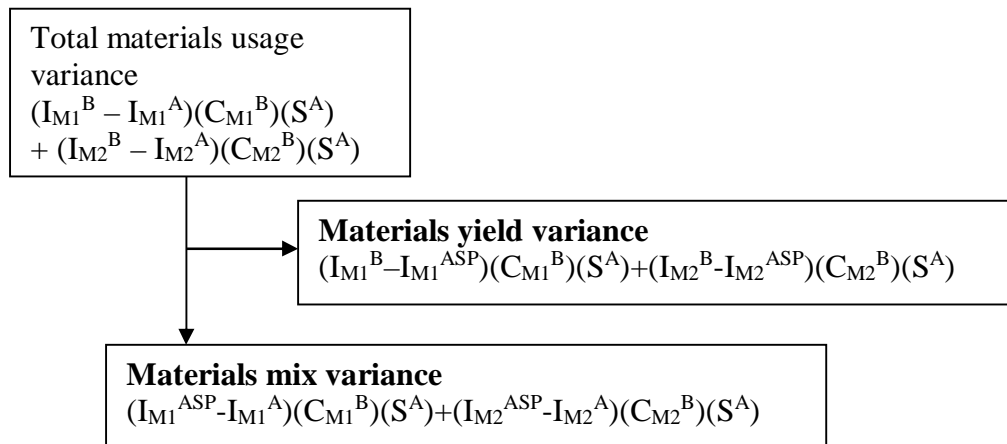
* Sales revenue volume variance
 = **sales growth variance** + **sales mix variance**



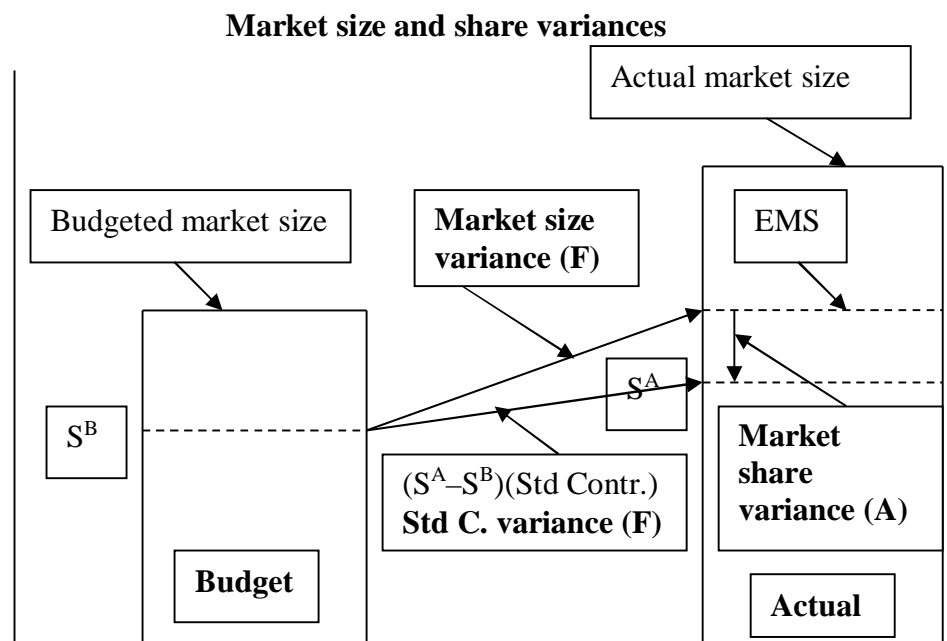
* Standard contribution variance
 = **market size variance** + **market share variance**



* Materials usage variance
= **materials mix variance** + **materials yield variance**



The importance of the mix and share variances to management can usefully be shown illustratively, and the example used is the market size and market share variance. Whilst the sales contribution variance might be favourable, this could be due to a favourable variance in terms of the growth in market size, but offset by an adverse variance in terms of market share. The market share is an important indicator of ‘competitiveness’, so if that is adverse, there could be problems arising, even in a growing market.



An example of the ‘flexed’ budget approach

The table shows the move from the budget figures, through ‘flexed’ budget figures, to the actual outturn figures. This distinction is:

Budget = Budgeted output at standard costs and prices

Flexed Budget = Actual output at standard costs and prices

Actual outturn = Actual output at actual costs, prices and input quantities

and this enables a matrix of variances to be prepared. Taking the economic costing basis first, we could have:

	Std per unit	Budget	Flexed Budget	Actual outturn
	£	S ^B =25	S ^A =20	S ^A =20
Revenue	2	50	40	42
<i>less</i>				
variable costs	1.2	30	24	28
<i>equals</i>				
contribution	0.8	20	16	14
<i>less</i>				
fixed costs	n/a	12	12	11.2
<i>equals</i>				
profit/loss	n/a	8	4	2.8

Given the fixed costs are 12, and the budgeted sales are 25, then the absorption method would show, given an absorption rate based on sales, of $12 \div 25 = 0.48$, and we can add these in for the n/a parts above, noting that with the absorption method the flexed budget shows a reduction (ie, under-recovery) of fixed costs because actual sales are lower than budgeted sales:

Contribution	0.8	20	16	14
<i>less</i>				
absorbed FC	0.48	12	9.6	11.2
<i>equals</i>				
profit/loss	0.32	8	6.4	2.8

Given these figures, the profit variance by either method is $2.8 - 8 = -5.2$ adverse. The variances can be calculated by taking the differences between the last three columns, and on the economic costing basis would show:

Selling price variance	42 - 40	= + 2	favourable
+ Std contribution variance	16 - 20	= - 4	adverse
+ Total variable cost variance	24 - 28	= - 4	adverse
+ <u>Fixed cost variance</u>	12 - 11.2	= 0.8	<u>favourable</u>
<i>equals</i> Total profit variance	2.8 - 8	= - 5.2	adverse

The standard contribution variance of - 4 could be split into the standard revenue variance $(40 - 50) = -10$ and the standard variable cost variance of $(30 - 24) = 6$, a total variance of - 4.

Using the absorption method there would be additional variances, given the standard contribution variance of -4 can be split into the standard profit variance $(20 - 25) \times 0.32 = -1.6$ and the fixed cost sales volume variance $(20 - 25) \times 0.48 = -2.4$, a total of -4.

A full decomposition of variances

The relationships between the variances (from left to right) can be shown as follows:

	Revenue(Q)	sales growth/mix	
Revenue(T)		market size/share	
	Revenue(P)		
Profit(T)		Std Contribution	Std Profit
	Costs(var)	Std variable cost	stock production
Costs(T)		Inputs price & efficiency	
	Costs(fixed)		FC over/under recovery

All of the variances covered can usefully be illustrated in a worked example chart, working from the top with the total profit variance to the bottom, with the arrows showing the splits between the various variances. The main variances are shown in bold. It is also shown whether the variance is from the budget to the flexed budget (B»F) or the budget to the actual outcome (B»A). Given the 'adding up' rule only a proportion of these variances would have to be calculated in order to complete the whole chart (any two will give a third). Data provided for the notations used in the formulae shown above can be programmed for the answer to be given automatically. The data for the variances shown in the chart came from the following data for two products (1 and 2), two inputs for materials (1 and 2), and the absorption rate for fixed costs was based on labour hours (L). The calculations for standard contribution, standard profit and the absorption rate are shown.

		Data table									
		<u>P</u>	<u>S</u>	<u>Q</u>	<u>I_{M1}</u>	<u>I_{M2}</u>	<u>Σ_M</u>	<u>C_{M1}</u>	<u>C_{M2}</u>	<u>I_L</u>	<u>C_L</u>
1	B	90	60	60	8	6	14	£2.25	£4.0	6	£5
	A	100	50	75	10	8	18	£2.0	£2.0	7	£6
2	B	60	100	100	2	10	12	£4.0	£1.6	5	£4
	A	50	130	120	2	10	12	£4.0	£1.6	5	£4

Additional workings

* Fixed cost budgeted = 516; Fixed cost actual = 540. Absorption rate based on budgeted labour hours is therefore $6 \times 60 + 5 \times 100 = 860$, total budgeted labour hours divided into budgeted fixed costs ($516 \div 860$) therefore equals the absorption rate per labour hour (r_L^B) = £0.6

*The standard contribution rate is revenue less variable costs, therefore for products 1 and 2 it is:

Product 1 = $90 - 72 = £18$, where 72 is the variable cost of labour ($6 \times £5$) plus the variable cost of materials 1 and 2 ($14 \times £3$), where £3 is the weighted average cost of materials 1 and 2 used to produce product 1.

Product 2 = $60 - 44 = £16$, where 44 is ($5 \times £4$) plus ($12 \times £2$).

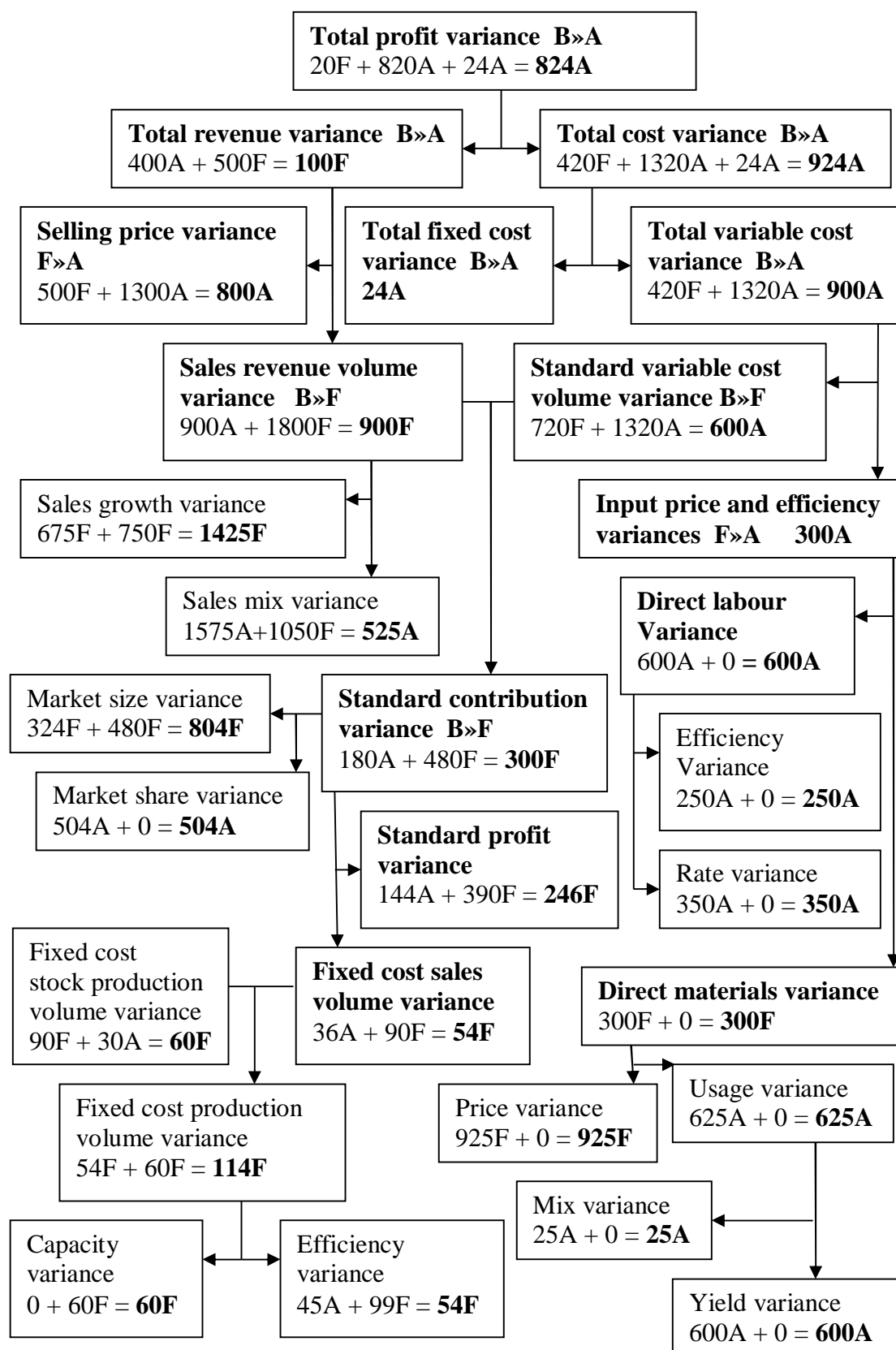
* The standard profit rate is standard contribution less absorbed fixed costs, therefore

Product 1 = $£18 - 6 \times £0.6 = £14.4$

Product 2 = $£16 - 5 \times £0.6 = £13.0$

The variances shown in the chart below derived from the figures above and the formulae shown are split between the variances arising from product 1 (put first) and product 2, plus an additional variance for the fixed cost expenditure variance where that forms part of the total variance, as with the total profit variance.

Decomposition chart



An illustrative worked example

Glamour Ltd manufactures made-to-order wedding dresses which are made from silk and a synthetic substitute. Its budget has to cover its fixed common costs of £240,000, which are recovered on labour hours, but there are opportunities to use the cheaper synthetic substitute rather than silk, although account has to be taken of the fact that it might affect demand. This year Glamour Ltd decided to reduce its price during the year by 14% but was surprised to find that although their sales increased by 50%, the market for wedding dresses as a whole went up by 75%. The data from Glamour's budget and outturn which is needed for its end of year profit and variance statement is as follows, with materials in square metres:

	Sales price	Sales quantity	Materials silk	Materials syn	Materials cost silk	Materials cost syn	Labour hrs	Labour rate	Fixed costs
Budget	£1000	1200	4	8	£40	£15	24	£10	£240,000
Actual	£860	1800	3	12	£55	£16	22	£11	£275,000

and its budgeted overhead recovery rate was $\text{£}240,000 \div 1200 \times 24 = \text{£}8\frac{1}{3}$ per labour hour. Glamour's standard contribution and standard profit per wedding dress was therefore:

	Price	1000
less	material	
	silk 4x40 =	160
	syn 8x15 =	<u>120</u>
		280
	labour	24x10 = <u>240</u>
		<u>520</u>
equals	standard contribution	480
less	recovered fixed costs	$24 \times 8\frac{1}{3} =$ <u>200</u>
equals	standard profit	<u>280</u>

and its **budgeted profit** would have been $\text{£}280 \times 1200 = \text{£}336,000$. Glamour's actual profit however was as follows:

Revenue	£860 x 1800	= 1,548,000
less Variable costs		
	Silk £55 x 3 x 1800	= 297,000
	Syn £16 x 12 x 1800	= 345,600
	Labour £11 x 22 x 1800	= 435,600
less Fixed costs		= <u>275,000</u>
equals Actual profit		= £194,800

compared with a budgeted profit of £336,000; a percentage reduction of 42% and a **total adverse variance** of **£141,200**. The variance statement used to explain this, using the formulae set out, would have been:

Variances	Favourable (+)	Adverse (-)
Selling price v		- 252,000
Standard contribution v	+ 288,000	of which: mkt size + 432,000
Materials price v		- 102,600 mkt share - <u>144,000</u>
Materials usage v		- <u>36,000</u> + <u>288,000</u>
of which:		- 138,600 and: standard profit+ 168,000
mix + 90,000		fixed cost abs + <u>120,000</u>
yield - 126,000		+ <u>288,000</u>
Labour efficiency v	+ 36,000	
Labour rate v		- 39,600
Fixed cost expenditure v		- 35,000
Total variances	+ 324,000	- 465,200 Total net variance - 141,200

and this can be used by the management of Glamour Ltd to learn some lessons from the year, in particular the adverse selling price variance combined with the falling market share, and the rise in their fixed costs.

Variance Reporting and Responsibility

The use of variance analyses should take account of:

- their '**economic relevance**' to decisions (eg, the accounting variances);
- hierarchy - the choice of main or '**core variances**' versus sub-detail (or 'of which' groups of variances, such as standard contribution variance, of which, market size variance plus market share variance)
- focus where possible on matching variances with those **responsible for 'causing' variations**;
- but reflect on whether the variance is a fair reflection of individual performance or causation (ie, responsibility implies traceable causation, and therefore distinguish **controllable** from **uncontrollable** outcomes);
- integrate variance analysis into a **learning culture** so that lessons are fed back to improve the control for future outcomes. A learning organisation has to decide whether to instigate a 'no-blame' culture in order to maximise learning opportunities and engagement, and when to apply disciplines in cases of clear, negligent fault. **In effect, the application of variance analysis has to be properly managed**, covering standard setting, incentives and penalties.

The problems of attributing responsibility can be readily identified in the following examples:

1. Was the standard cost reasonably set by the management in the first place (ie, was it reasonably achievable)?
2. Has the standard which was achievable changed in the meantime due to uncontrollable external events (eg, a world price rise due to an unforeseen shortage of certain materials)?
3. Were the variances independent or interrelated? Hence, if sales are an 'inverse' function of price, then a favourable price variance from a rising price will be matched by an unfavourable standard contribution variance from the quantity of sales going down. In this case both have to be considered together (a joint variance). Another example is where a buyer has achieved lower prices on inputs of materials but at the expense of quality, which has adverse affects on usage. As before, the materials price and usage variances would in part have to be considered together. The problem is to decide what proportion of the change is contributed to by the interaction between the variances. Finally, where an input variation has an impact on the total output achieved (because there are a limited number of labour or machine hours), then the consequential loss of standard contribution has to be included as well.

Clearly, a thorough analysis designed to attribute responsibility fairly, and to aid 'feed-back' control, may involve management in some work! This is why text books and academics traditionally focus on three causes of variance:

- incorrect original plans or standards;
- inefficiency in operation;
- change in environmental circumstances.

The result is to distinguish '**appraisal variances**' for attributing responsibility to individuals and '**control of planning variances**'. The two are integrated by utilising the concepts of controllable and uncontrollable outcomes, and perhaps by formally incorporating 'revised' standards based on changing environmental circumstances into the formal appraisal variance (hence, for example, C_M^B could be revised to C_M^{RB} for comparison with C_M^A if the purchasing manager was to be held responsible for an adverse (or favourable) materials cost variation).

Note:

- The fixed cost expenditure variance should be analysed in some detail (given it is the 'sink' to which many uncontrolled costs find their way from headquarters!).
- Investment decisions should be subject to post-investment variance appraisal.

Control of variances might be facilitated by a combination of 'carrots and sticks', and modern methods of remunerating 'effort' are designed to achieve this, as far as they can, whilst noting that money is not everything in motivation. Examples include the move from standard annual salaries to the following in some proportion or other:

- basic salaries and wages (on-going unless terminated, but note too the motivation of expectations of progression from good current performance)
- bonus schemes (related to current and past performance)
- share option schemes (share price as a measure of future expectations about performance)

Tracing responsibility for performance is often complex, and not necessarily overcome by changing the remuneration scheme. For example, are share options benchmarked to relate to variations from the stock exchange index as a whole?

2. Control by accounting standards, regulatory accounts and auditing

If economics is about the effective allocation of scarce resources, accounting is about recording the value of those scarce resources. Accounting therefore plays a central role in all economies, and this role is underpinned by 'normative' economics which sets the conceptual framework for regulating the calculation and use of accounting information.

Accounting information is the network which holds together the parties to economic transactions, and influences the decisions which they might make. It is therefore clearly vital that accounting information should have various essential properties, such as accuracy, coherence and economic relevance. Reliability is also an essential property because all economies operate on a separation of roles and responsibilities, and these roles have a 'hierarchy' in relation to control over information. Accounting information which passes between parties therefore has to be regulated to ensure that one party is protected from abuse by another, and to promote the public interest. From this need comes accounting standards and auditing.

The primary model used to describe the separation of roles is the 'Principal/Agent model, which reflects the separation of ownership from management control in the modern economy. The 'agent' should act on behalf of the 'principal', and serve the principal's interests, but a problem arises where the agent is fundamentally in control of the outcomes, and shapes them to serve the agent's interest. The idea of the principal/agent model is found in both the private and public sectors:

* Companies produce goods and services which consumers can buy, but the companies are run by directors who control and manage the company, although the company is owned by its shareholders. The shareholders are the principals whilst the company directors are the agents.

* The democratically elected central or local governments produce goods and services which are financed by tax revenue or sales, and so the citizens/voters are the principals whilst ministers, councillors and government institutions are the agents.

The first step to ensure that the interests of the principals are protected is by requiring the financial outcomes of agents' decisions to be audited. Audit should be carried out by organisations independent of the agents and appointed by the shareholders, and therefore the

audit process is meant to represent the interests of the principals, not the agents. Audit of private sector companies is carried out by professional audit firms, having been appointed by the shareholders at the annual general meeting. This type of arrangement has been criticised because the board of a company recommends the auditor to the shareholders, and the relationship of the audit firm to the agents may be influenced by the fact that audit firms can receive non-audit contracts from the company, including reviews of financial strategy. Audit of state organisations is carried out by organisations such as the National Audit Office (NAO) or, in local government by the District Audit Service and the Audit Commission.

The normative aspects of auditing can be seen in the audit statement which accompanies a company's annual report and accounts. A 'clean' audit certificate would simply say that "these accounts give a 'true and fair' view of financial position and profit or loss", and that the accounts comply with the appropriate 'accounting standards'. The true and fair concept is important because accounts must not only be accurate, but meaningful, particularly in terms of whether the company can be considered to be a 'going concern'. Such a concept is important when considering the banking crises which have emerged since 2008, because if the directors of banks (the bankers) had been responsible for reckless financial behaviour, it should have been the job of the auditors to draw attention to the problem. The 'qualified' audit certificate is the means by which shareholders and other interested parties, notably creditors, are given the information which allows action to be taken.

Creative accounting

The second step is the need for financial information to be 'standardised' by setting accounting standards which should be complied with both professionally and legally. By ensuring that financial statements are prepared in accordance with 'generally accepted accounting practice' (GAAP), it is both easier to rely on the accounts and to compare the results with other companies and comparators. The role of accounting standards is therefore to reduce the range of alternatives available in preparing accounts, and to focus on making accounting information economically useful.

The need for accounting standards, and the process by which they can be established, was well illustrated by Ian Griffiths in his 1986 book **'Creative Accounting – How to make Profits what you want them to be' (2 + 2 = 5)**. His opening paragraph sets the tone: "Every company in the country is fiddling its profits. Every set of published accounts is based on books which have been gently cooked or completely roasted. The figures which are fed twice a year to the investing public have all been changed in order to protect the guilty. It is the biggest con trick since the Trojan horse". He goes on to say that without effective accounting standards, such activities can be considered to be totally legitimate – "It is creative accounting". In this way "above-the-board means of achieving underhand ends is rife in boardrooms throughout the country as companies contrive to translate their activities of the year into reported results which flatter the management and the share price".

However, in terms of the principal/agent model, it has to be recognised that it is not just a question of a company's directors' interests versus its shareholders interests. There are situations where a company's directors and shareholders have a common interest in 'misleading' others. The possibility of creative accounting arises because accounts are more than just a cash flow statement. Fixed assets in the balance sheet can be revalued at the stroke of a pen (dr Fixed assets, cr revaluation reserve); stock can be valued on various bases (LIFO, FIFO or average cost), each method resulting in a different calculation of profit. Future liabilities can be provided for in advance or ignored; and certain transactions might be treated as 'balance sheet movements', rather than passing through the profit and loss account. The subject obviously

becomes a complex one, but the general tenor and scope of the debate can be seen from the history of standard setting in the UK since around 1970.

Standard setting

In the UK, prior to 1970, the control of accounting practice was by professional self-regulation. There were three separate bodies of chartered accountants, one each for England and Wales (ICAEW), Scotland and Ireland which, along with certified accountants (ACA), focused on financial accounting and auditing, whilst CIMA, the cost and management accountants, focused on management accounting, and the Chartered Institute of Public Finance and Accountancy (CIPFA) focused on the public sector, particularly local government. The six accountancy bodies were brought together by the Consultative Committee of Accountancy Bodies, the CCAB.

However, financial scandals and corporate collapses undermined confidence in self-regulation, and the first step in reform was to create an Accounting Standards Committee (ASC), which issued Statements of Standard Accounting Practice (SSAPs), and less obligatory statements, the Statements of Recommended Practice (SORPs). The ASC issued a large number of accounting standards in the 1970s and it soon became evident that to receive a 'clean' audit certificate, preparers of accounts would need to comply with accounting standards, unless it could be shown that a true and fair view required an 'override' which would allow the accounts not to be prepared in compliance with all the accounting standards. If used, the reasoning has to be disclosed and the difference between the accounting standards compliant accounts and the 'true and fair' accounts shown. Subsequent Companies acts made it a legal requirement to comply with accounting standards.

The ASC's most controversial standard was SSAP 16, Current Cost Accounting, issued in 1980 following the exposure draft published in 1976. The sharp increase in inflation during the 1970s meant that the difference between the historical cost of company assets and their current financial worth separated markedly. The debate centred on the fact that historical accounting was 'accurate', but that current cost accounting gave economically 'relevant' financial information. One problem for achieving consensus was that some companies preferred historical cost accounting because it kept the depreciation charges lower, and the accounted profit higher, along with a higher return on capital employed (ROCE) ratio. However, current cost accounting received both strong academic and government support in terms of the principle of 'capital maintenance', and particularly in the context of regulating the newly privatised utilities during the 1980s. A notable contribution was from Professor Bryan Carsberg, an accounting professor at the London School of Economics (LSE), and who became the first Director General of OFTEL following the privatisation of British Telecommunications (BT) in 1984, when the ASC published his 'The Usefulness of Current Cost Accounting' in 1982. The debate continued however, and when inflation began to fall SSAP16 was withdrawn, allowing more choice in preparing accounts on capital maintenance concepts.

One problem for the ASC was that it had been formed from the professional bodies, overseen by the CCAB, and the majority of the members (15) represented 'preparers' of accounts, whilst 'users of accounts' were only represented by 5 members. By the end of the 1980s, the demand for a more 'user' orientated organisation led to the formation of a government/profession funded Financial Reporting Council (FRC) and Financial Reporting Standards Board (known as the ASB) to replace the ASC. Financial Reporting Standards (FRSs) replaced the SSAPs, although SSAPs in place were retained until they were superseded by a new FRS, if necessary. The FRC treated the accounts as part of financial reporting as a whole, where the notes to the accounts are as important as the accounts themselves, and have to be read alongside the annual report of the

company to its shareholders and other interested parties, which would include an Operating and Financial Review. An important distinction was drawn between:

*User protection – ‘disclosure’ requirements to ensure that preparers of accounts cannot abuse their ‘monopoly’ power over the provision of information.

*Market efficiency – ‘measurement’ requirements to help improve the efficiency of markets by giving relevant information.

Financial Reporting Standards

The role of the ASB is to issue accounting standards, and was recognised in the Companies Act 1985, taking over the task from the ASC in 1990. By this time multi-national companies were an important part of the ‘global’ economy, and the ASB had to collaborate with standard setters in other countries, as well as the International Accounting Standards Board (the IASB). Where possible in terms of the legal framework in each country, national standards should be aligned with international standards.

The names of the accounting standards currently in place for the UK (2011) set the tone for the range of issues and complexity arising in preparing accounts, and the emphasis on both disclosure and measurement. First, the SSAPs still retained by the ASB:

- SSAP 4 Accounting for Government Grants
- SSAP 5 Accounting for Value Added Tax
- SSAP 9 Stocks and Long-term Contracts
- SSAP 13 Accounting for Research and Development
- SSAP 17 Accounting for Post Balance Sheet Events
- SSAP 19 Accounting for Investment Properties
- SSAP 20 Foreign Currency Translation
- SSAP 21 Accounting for Leases and Hire Purchase Contracts
- SSAP 24 Accounting for Pension Costs
- SSAP 25 Segmental reporting

The financial reporting standards are a longer list and are often compliant with international accounting standards:

- FRS 1 Cash Flow Statements
- FRS 2 Accounting for Subsidiary Undertakings
- FRS 3 Reporting Financial Performance
- FRS 4 Capital Instruments
- FRS 5 Reporting the Substance of Transactions
- FRS 6 Acquisitions and Mergers
- FRS 7 Fair Values in Acquisition Accounting
- FRS 8 Related Party Disclosures
- FRS 9 Associates and Joint Ventures
- FRS 10 Goodwill and Intangible Assets
- FRS 11 Impairment of Fixed Assets and Goodwill
- FRS 12 Provisions, Contingent Liabilities and Contingent Assets
- FRS 13 Derivatives and Other Financial Instruments: Disclosures
- FRS 14 Earnings per Share
- FRS 15 Tangible Fixed Assets
- FRS 16 Current Tax
- FRS 17 Retirement Benefits
- FRS 18 Accounting Policies
- FRS 19 Deferred Tax
- FRS 20 (IFRS2) Share-based Payment
- FRS 21 (IAS10) Events after the Balance Sheet Date
- FRS 22 (IAS33) Earnings per Share
- FRS 23 (IAS21) The Effects of Changes in Foreign Exchange Rates
- FRS 25 (IAS32) Financial Instruments: Presentation
- FRS 26 (IAS39) Financial Instruments: Recognition and Measurement
- FRS 27 Life Assurance
- FRS 28 Corresponding Amounts
- FRS 29 (IFRS7) Financial Instruments: Disclosures
- FRS 30 Heritage Assets

Out of such a long list of accounting standards, two accounting standards in particular set the context for the use of accounting standards. First is FRS 18, Accounting Policies, which requires preparers of accounts to tell users of accounts which accounting policies have been used to prepare the accounts, and to show that these are consistent with the accounting policies used in the comparative accounts from previous years. Second is FRS 3, Reporting Financial Performance, which integrated historical cost accounting with 'modified' historical cost accounting through 'The Statement of Total Gains and Losses'. It also separated 'continuing' from 'discontinued' operations in order to identify the underlying financial trends, and clarified the use of 'exceptional' and 'extraordinary' items in reporting accounted profit. In effect ensuring that so-called exceptional items were included.

Two other important statements published by the ASB are:

- * **The Operating and Financial Review** – which provides a framework for company directors to set out the business performance and factors affecting the results in their annual report.

- * **The Statement of Principles for Financial Reporting** – which sets out the objectives of financial reporting and defines the qualitative characteristics of financial information, being Relevance, Reliability, Comparability, Understandability and materiality. It also sets out the basis for 'recognising' financial transactions and the principles for effective presentation in annual reports and accounts.

The relationship of the UK's accounting standards to the international accounting standards issued by the IASB continues to develop because the needs of the global economy really require multi-national companies to be able to list on a number of different countries stock exchanges and not have to restate their accounts in different countries to comply with each country's own standards and legal requirements. If UK GAAP were to show a profit, but US GAAP show a loss for the last year, then whilst it might be useful for the specialist to understand the reason for the difference, it would certainly be a problem for making accounts understandable and maintaining investor confidence. Inevitably, the process of international convergence takes a long time. In part it relates to many countries' belief that their standards are the best one, but also different legal frameworks, differences of opinion as to whether the standards should be based on 'principles' or set out complex lists of requirements, whether accounting standards are firm enough in restricting the possible accounting choices available, and whether accounting standards need to be incorporated into the law of the country in detail. Developing countries may also take a different position from the developed countries with their own accounting institutions and tradition, and one aim of the IASB is to take account of the financial reporting needs of emerging economies. In spite of the difficulties, the general trend is towards accepting accounts compiled in accordance with international financial reporting standards. Much depends on getting political support, as has been the case in the EU in accepting European Public Sector Accounting Standards, and the International Federation of Accountants (IFAC) continues to play an important role in promoting global standards.

Regulatory accounts

Accounting standards are a method of regulating the preparation of annual reports and accounts, but the term regulatory accounts came to the fore in the UK with the privatisation of the utility and network industries, and their regulation. A regulated utility might wish to list on the stock exchange using historical costs, while the regulator might wish to set price controls based on current cost accounting. In addition, the regulated utility might only be a 'subsidiary' of a larger 'parent' company. Inevitably many companies had to publish two sets of accounts, and there was concern about the burden which the information requirements of the regulators imposed. To set this burden in context, a report was published in 2001 entitled 'The Role of Regulatory Accounts in Regulated Industries', prepared by the chief executives and director

generals of Ofgem (the office of gas and electricity markets), Oftel (the office of telecommunications – now Ofcom), Ofwat (the office of water services), ORR (the office of rail regulation), CAA (the civil aviation authority), the Postal Services Commission, and Electricity and Gas Supply – Northern Ireland.

The development of regulatory accounts has made a large contribution to academic and policy debate, in part because of the diversity of the regulated industries. This was revealed by the report when it said in a section on the ‘common regulatory accounting framework’: “Regulatory accounts should be prepared and audited using either the Regulatory Accounts Guidelines (RAGs) for that industry or, where a RAG does not cover the issue, UK GAAP. Where there is any conflict between RAGs and UK GAAP, then the RAGs will take precedence”. The report went on to say, “The approach adopted by each regulator to the basis of preparation reflects the specific circumstances of the industry concerned. Given this diversity of valuation methods it is not possible to achieve consistency in the basis of preparation of regulatory accounts and each regulator will use the method that is most appropriate for its industry”. It can be said that the ‘accountability’ of both the regulated industries and their regulators has become clearer given the development and debate on regulatory accounts.

The public sector and whole of government accounts (WGAs)

The UK ‘public sector’ is a large conglomerate of many different types of institution, ranging from central government departments, organisations attached to those departments (such as the military, the police service, and the health service), local governments, agencies (such as the environment agency), and a variety of other organisations. The development of accounting standards, and the high level of public debate concerning such things as the size of the public sector compared with the private sector, and the size of the national debt, meant that there was an increasing focus on the way public sector accounting was practised.

Parliament is required to hold the government to account, and each year the process involves debate on the public expenditure proposals of the government, and its tax intentions, generally known as the ‘budget’ cycle. Parliament approves the ‘supply’ of cash and the National Audit Office audits the accounts to state whether what has been spent is in accordance with what has been approved. The Public Accounts Committee then produces its own report. The ‘cash-based’ approach is a system which can work well because it is simple, and understandable. Revenue comes in from general taxes (such as income tax), specific taxes (such as VAT or car tax) and income from the sale of goods and services. Expenditure goes out to pay salaries, contracted work and supplies of goods and materials. If revenue exceeds expenditure, then net borrowing can go down, but if not net borrowing has to increase and the public debt goes up.

The problem is that it is unclear whether the balance of revenue to expenditure each year is ‘correct’ from an accounting perspective. Tax revenue is treated as ‘unhypothecated’ so it cannot be matched with, or reserved for, any particular expenditure. For example, car tax is not considered to be the fund to finance road maintenance. In company accounts, a distinction is made between operating and capital spend, where operating expenditure has to be charged as an expense to this year’s accounts, whilst the capital expenditure can be spread over many years, the only expense each year being the ‘depreciation’ charge. Revenue linked to these expenses then determines whether a profit (surplus) or loss (deficit) is being made.

By 2000 it had been decided that the public sector should have an ‘accruals’ based accounting system, and that this system should bring together all the public sector institutions into one ‘consolidated’ set of accounts. From this, the public would be more aware of the assets the state owns, what liabilities it faces looking forward, and whether operating expenditure is at

least being covered by current revenue, rather than relying on more borrowing. The process started in 1997/8 when the Treasury carried out a ‘scoping’ study and concluded that a consolidated ‘whole of government’ set of accounts was the best way to proceed. A commitment was made in the Code for Fiscal Stability, and this became a statutory requirement in the Government Resources and Accounts Act 2000.

It took ten years until a WGA could be published in 2011 for the year 2009/2010. This is not surprising given the complexity of the task. The consolidated accounts had to bring together the audited accounts of central government departments, local authorities, devolved administrations, the national health service, and public corporations, some 1500 bodies in total. Given that their accounts might be prepared on different accounting principles, it meant that adjustments would have to be made to bring them onto the same basis, and to eliminate any ‘intra-government’ transactions. The Treasury announced that the 2009/2010 accounts were based on the EU adopted international financial reporting standards, adapted or interpreted for the public sector context, and were complements to the national accounts figures produced by the Office for National Statistics. The WGAs were similar in presentation to private sector accounts. In the financial reporting ‘manual’ for WGA the Treasury set out the accounting standards applied, and on publication in July 2011, the Chancellor of the Exchequer said that it “represented a step change in transparency”.

The Comptroller and Auditor General presented his audit report to the Public Accounts Committee in October 2011, covering the following financial statements and related notes and annexes:

- * Consolidated Statement of Revenue and Expenditure
- * Consolidated Statement of Financial Position
- * Consolidated Statement of Changes in Taxpayers’ Equity
- * Consolidated Cash Flow Statement

The accounts were qualified for a number of reasons, which might have been expected given the complexity of the task, and given this was the first set of WGA.

The qualifications related to:

- * Exclusion of some ‘public bodies’ from the WGA ‘boundary’. The Treasury had based those to be included on those classified as public bodies by the Office of National Statistics. In particular the exclusion of NetworkRail, publicly owned banks, and the Bank of England was identified as of particular concern for the Statement of Financial Position.
- * Inconsistent application of accounting policies. A number of bodies consolidated into the WGA do not adopt the same accounting framework under which the WGA is prepared. In particular, bodies in the local government sector follow the Local Government Statement of Recommended Practice (SORP) for 2009/10, which is based on UK GAAP.
- * Lack of evidence that elimination of intra-government transactions and balances had been achieved.
- * Inclusion in the WGA of bodies whose own accounts had been qualified, noting in particular that the Ministry of Defence’s resource accounts had been qualified on two grounds which were of material significance to the WGA.

The WGAs were qualified for the second consecutive year in 2012, concerning such things as the valuation basis used for local authority assets on an historic cost basis, and the treatment of Network Rail. The major problem for the WGAs is probably not the qualifications but that fact that they are not being widely used or referred to. They were not mentioned in the

government's Budget 2013 documents, and the Public Accounts Committee has asked the Treasury to increase the use of WGAs in Whitehall.

But overall the WGA was considered to be a true and fair view. The National Audit Office carries on with an important role in the political economy of the UK which began in 1314 with the appointment of the first Comptroller General of the Exchequer, and the Exchequer and Audit Department was formally established in 1866. The National Audit Office was created in 1983 by the National Audit Act and as it says in its 'guide to the National Audit office' the "primary concern is accountability to Parliament, and ultimately the taxpayer – to assure them that public funds and resources are used properly and to good effect".

The reference to good effect reflects the extension of the National Audit Office's role into 'value for money' investigations, and these may result in up to 50 reports a year. It is important to note that in carrying out this role the NAO is only meant to focus on ways in which central government can improve their management and increase cost-effectiveness, it is not about questioning the policies of a democratically elected government. However, deciding on where that line should be is often a difficult one!

III Analytical Examples

Having set out the foundations for economic analysis and accountancy, the next step is to consider various analytical examples which can be considered a little more difficult, but utilise the underlying principles already set out in the foundations.

1. Reconciling NPV and IRR decision rules

The standard investment rules have been set out in the foundations, and these rules are covered in practically all economic and accountancy textbooks. The problem is that they often end there. This means that standard questions are often not easily answered. Let us take a simple range of investment combinations: ‘independent’ projects with, or without, capital rationing; and ‘mutually exclusive’ projects, again with, or without, capital rationing. Independent projects are those which you can invest in as many of them as required; mutually exclusive projects are those where one or the other is chosen, such as choosing to build a tunnel or a bridge over a river; and capital rationing is where there is insufficient money to invest in all the worthwhile projects available, hence you have to prioritise.

These combinations can be set out in a two by two matrix, which we call the ‘window’ on investment appraisal. The next step is to ask to which of the four ‘panes’ in the window do the standard investment rules clearly apply. The answer will usually be given as shown in the chart, but clearly that leaves one pane empty, and only one pane with the IRR included. The next question is then, how should the rest be filled in? The answer to that is the need to calculate ‘modified’ IRRs which are consistent with NPV decision rules.

The Window on Investment Appraisal

	No capital rationing (NCR)	Capital rationing (CR)
Independent projects (IP)	Invest if: $NPV > 0$ $IRR > r^*$	Rank projects on NPV/C for inclusion in the investment list
Mutually exclusive projects (ME)	Choose the project with the highest NPV	

The IRR is the first problem to be addressed because it is often asked whether the IRR can be used to rank independent projects subject to capital rationing (IP/CR), or to choose mutually exclusive projects not subject to capital rationing (ME/NCR)? Text books often make it no easier by saying that the IRR may anyway be faulty because in cash flow series which are not strictly a ‘normal’ series of an investment followed by a series of receipts ((x), x, x, x,.....), there are multiple solutions for the IRR. The answer tends to be, ignore the IRR as a decision rule and simply use NPV decision rules. Unfortunately this is not necessarily a good answer, as

people are used to the idea of the rate of return, and you might be asked for an IRR analysis as well as the NPV analysis. In practice, cash flow series, even with changes of sign, do not often result in meaningful multiple IRRs but when they do the problem is that none of them necessarily have any economic relevance, unlike r^* and the NPV. For example, three IRRs will result from the following cash flow series, -102 in year 0; +200 in year 2; -103 in year 5; and +5 in year 25. There are three changes of sign in the series and the resulting IRRs are 0%; 2.15% and 8.81%. Apart from being a rather unrealistic cash flow series for most business situations, the main point is that it is more useful if there can be a 'modified' IRR which is a single IRR and economically relevant like r^* and NPV

To calculate appropriate modifications we need to consider the 'normal' business range, being rates of interest between -100% and +100%. This range helps to show why multiple IRRs can occur. The next step is to calculate a modified IRR for each of the three panes which have been filled in. The modified IRRs will be as follows:

For IP/NCR: The Investment IRR (**IIRR**) and the Terminal IRR (**TIRR**)

For IP/CR: The terminal IRR with a common terminal year (**TIRRC**)

For ME/NCR: The standardised TIRRC, being the **STIRRC**

The normal business range and multiple IRRs

This is of practical importance because it shows why multiple solutions to the IRR occur. For positive rates of interest we have 'discount' factors, but for negative rates of interest we have 'multipliers' because $1/(1-r)^t$ will be greater than 1. A 'normal' investment (I) NPV curve of the structural form (I), x, x, x,.... will descend from + infinity at minus 100% r but a non-normal cash flow series of the form (I), x, (x) will have its NPV curve rise from minus infinity at minus 100% r . Each will tend towards I as the rate of interest approaches +100%. The affect of this can be shown in a chart, but it is also useful to show the present value of receipts curve as well as the present value of payments curve, given $PVR-PVP = NPV$.

Comparing the range of discount factors between - 100% r and + 100% r , for various rates of interest and number of years t , is useful to understanding why the PVR and PVP curves can cross over each other and hence cause the NPV curve to equal 0 at that point.

% interest rate	-100%	-50%	-10%	0%	+10%	+50%	+100%
$r =$	- 1	- 0.5	- 0.1	0	+0.1	+0.5	+1
$(1 + r) =$	0	+ 0.5	+0.9	1	+1.1	+1.5	+2
$1/(1+r)^t =$							
for $t = \text{now}$	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>
for $t = 2$	∞	4		1	.826	.444	.25
for $t = 5$	∞	32		1	.621	.132	.031
for $t = 10$	∞	1024		1	.386	.017	.00097
for $t = 30$.057		

It clearly shows that given higher rates of interest, or a longer period of time until a cash sum is receivable or payable, it will discount its value to a very small figure. The shape of the various curves will be very much affected by the sequence of positive and negative numbers in the cash flow series. A simple example can illustrate this, a payment now and at the end of year 2, with a receipt at the end of year 1.

0	1	2	Net cash flow
(200)	510	(315)	-5

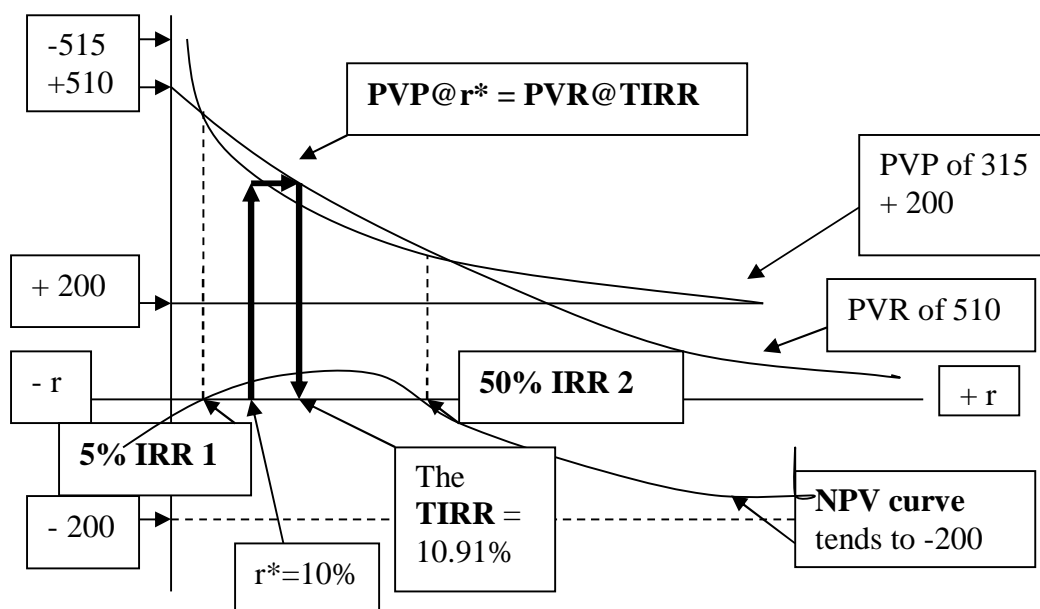
The question is whether, as a non-normal cash flow series, you would expect it to have one or two IRRs (or none) in the normal business range? A simple test would be to calculate the NPV at various interest rates:

Interest rate	NPV
0%	$510 - 200 - 315 = -5$
10%	$510(.909) + (-315).826 + (-200) = +3.4$
30%	$510(.769) + (-315).592 + (-200) = +5.71$

It is important to note that even for cash flows which when added up show a negative cash flow (- 5), they can still have a positive NPV at higher interest rates, in this case because the final cash payment (- 315) is discounted more than the receipt (+ 510), received a year earlier. The calculations clearly show that the NPV has been zero at an interest rate less than 10%, but we know too that the NPV curve must become negative again because at higher interest rates the 510 and 315 will be almost discounted away, leaving only minus 200. Given this simple cash flow, the two IRRs can be directly solved for by using a quadratic equation. Given we can set $-200 + 510(df1) - 315(df2) = 0$, if we multiply through by $(1+r)^2$ and then simplify we arrive at the quadratic equation $-200r^2 + 110r - 5 = 0$. This quadratic has two solutions for r where $r = 5\%$ and $r = 50\%$, which is in line with the shape of the NPV curve we were expecting.

The chart is drawn to illustrate this cash flow example, and will also be useful for showing the TIRR set out in the following section. The PVR and the PVP curves are shown on the same vertical axis, whilst the NPV curve is shown with the normal plus and minus numbers shown separately on the vertical axis. This is to show the important intersection (cross-over) points clearly.

Multiple IRRs and the Terminal IRR (TIRR)



The IIRR and the TIRR

The modification to achieve a 'single' IRR which is consistent with the NPV decision rule reflects the purpose of having an economically useful IRR, that is, the rate of return which is meaningful in terms of return on investment. In this case we can either say that the present value of our payments, discounted at the cost of capital to the firm (r^*) is the 'investment' (hence the investment IRR, being IIRR), and the receipts are the return on that, or we can say

that the investment is the original sum invested now (the 200), and the return is the future value of the remaining cash flows (both receipts and payments) at the end of the cash flow series (its 'terminal' year, hence the TIRR). Each element of the cash flow has to be compounded forward to the terminal year as 'reinvestment' at the cost of capital. In this case, if the cost of capital was $r^* = 10\%$, the calculations would be:

The **IIRR** is set $200 + 315(1/(1+r^*)^2) = 510(1/(1+\text{IIRR})^1)$ and solve for IIRR.

The **TIRR** is set $200 = \{510(1+r^*)^1 - 315\}(1/(1+\text{TIRR})^2)$ and solve for TIRR

The calculation for IIRR would be $200 + 315(.826) = 460.19$ and set $460.19 = 510/(1+\text{IIRR})^1$ so that $\text{IIRR} = 510/460.19 - 1 = 10.82\%$. In this case the IIRR is greater than the cost of capital, which is consistent with the NPV rule that if the NPV is positive at the cost of capital, then invest. The relationship between r^* and the TIRR in this case is shown on the chart as well. The calculation for the TIRR would be $510(1.1) = 561$, its future value at the end of the terminal year. Added to the payment in the terminal year of (- 315), the net future value would be 246. By setting its discounted value equal to the initial investment we can solve for the TIRR, which would be $200 = 246(1/(1+\text{TIRR})^2)$, therefore $\text{TIRR} = \sqrt{246/200} - 1 = 10.905\%$. The IIRR and the TIRR give slightly different results for their modified IRR because their chosen investment base is different, but both are consistent with the standard investment rules.

Ranking IP/CR projects with the TIRRC

The compounding forward method of calculating the TIRR can be used for ranking independent projects subject to capital rationing, but there has to be a common terminal year for all the projects to be ranked, hence TIRRC. This is to make sure that like is being compared with like. The standard investment rule is to calculate NPV/C for each project and rank them from the highest to the lowest. The highest is the highest £s per £1 of capital spend, so we choose to invest in projects from the highest downwards until the capital budget available is exhausted. The purpose of the TIRRC is to achieve the same, but this time using % points of TIRRC per £1 of capital spend.

In some textbooks the TIRR is referred to as the 'average' or 'extended yield' IRR. This is an important description because it is necessary for the calculation of the TIRRC to 'reinvest' the cash flows as they are realised until the terminal year at the cost of capital r^* , not the IRR. The idea of average comes from the fact that the TIRRC is a weighted average of combinations of returns (IRR and R^*) for each element of the cash flow series. It is quite natural to focus on the average IRR, representing the return over the life of the project, rather than the traditional IRR which focuses on returns for each element of the cash flow, even though they are for different periods of time. This can be illustrated from an investment cash flow series as follows (noting that the sum of the present values for each element of the cash flow series discounted at the IRR equals the initial capital outlay (ie $(x) = \sum(x.\text{dfs})$):

	0	1	2	3	Future Value at year 3
Cash flow series	(x)	x1	x2	x3	
x1 return@ IRR for 1yr		x1	then @ r^* for 2yrs		= FV x1
x2 return@ IRR for 2yrs			x2	then @ r^* for 1yr	= FV x2
<u>x3</u> return@ IRR for 3yrs				x3	= <u>FVx3</u>
= (x)					\sum FVs

If the cash flow series was (200), 100, 100, 100, for a cost of capital of 10% then the sum of future values would be $100(1.1)^2 + 100(1.1)^1 + 100 = 331$ and the TIRRC would be where $200 = 331(1/(1 + \text{TIRRC})^3)$. The TIRRC would be $= \sqrt[3]{331/200}$. Given that we have a future value of

331, and an investment of 200, then $200/331 = 0.6043$ would be the discount factor for a year 3 receipt which makes the NPV zero. Look in the discount tables for the rate of interest where a third year receipt would be valued at approximately 0.6043. This is approximately 18%. Given this cash flow series is an annuity of 100 per year, then the traditional IRR can be found from the annuity tables because we are setting $200 = 100(\sum \text{dfs})$. The sum of the discount factors is equal to 2 for this three year annuity, therefore the IRR is approximately 23%. The TIRRC is therefore 18%, being a weighted average of returns at $r^* = 10\%$ and $\text{IRR} = 23\%$.

Choosing ME/NCR projects using STIRRC

The final step in filling in the first three panes of the window on investment appraisal is to consider how the IRR should be modified for mutually exclusive projects with no capital rationing. The answer once again concerns measuring like with like. As with TIRRC we need a common terminal year, but we also need in this case a common level of investment for each mutually exclusive project being compared. The answer is to take the most expensive project in the list and make sure that all the other projects being compared with it spend the same amount. Given two mutually exclusive projects A and B, and where A costs 250 and B 150, then we can make their investment sum equal by saying B involves investing 150, with 100 remaining to be invested elsewhere at the cost of capital r^* . If for project B its TIRRC was 10%, and the cost of capital 5% then the standardised TIRRC would be the simple weighted average:

	<u>£m</u>	<u>proportion</u>	<u>% return</u>
Investment in B at TIRRC	150	60%	10%
Residual investment at r^*	<u>100</u>	<u>40%</u>	<u>5%</u>
Total 'common' investment	250	100%	8% = STIRRC

The mutually exclusive investment with the highest NPV will be the one with the highest STIRRC. The simple weighted average approach is a good approximation but could be inconsistent with the NPV rule because ideally the weights should be calculated using not the proportions based on the investment sums at year 0, but on the future value of those sums at the terminal year, otherwise it is a 'biased' estimate. The correct weights would therefore be calculated as:

		proportion
Future value of investment in B	$= 150(1 + \text{TIRRC})^t = a$	$a/a+b$
Future value of residual investment	$= \frac{100(1 + r^*)^t}{= b}$	$b/a+b$
Total future value	$= \sum = a + b$	

The window on investment appraisal can now be more fully filled in. It should be noted that the IIRR equals the standard IRR for a normal investment cash flow series, as that too has a single result. We have shown how the IRR can be modified to make it consistent with NPV decision rules, but two further points should be made before we consider the last pane. Some textbooks suggest that the IRR can be used to make 'pair-wise' comparisons between mutually exclusive projects not subject to capital rationing as long as the choice is tested for consistency using the 'incremental IRR' between the two projects, and amending your choice if this proves it to have been inconsistent. There is also the question of whether it is useful to divide the TIRRC into two parts: the compound part related to the cost of capital r^* and the part reflecting the underlying annual return related to the NPV (here to be called the 'z' factor).

The Window on Investment Appraisal with 'modified' IRRs

	NCR	CR
IP	NPV > 0 IRR or TIRR > r^*	NPV/C ranking TIRRC ranking
ME	NPV = highest STIRRC = highest	?

Incremental analysis consistency test for ME/NCR projects

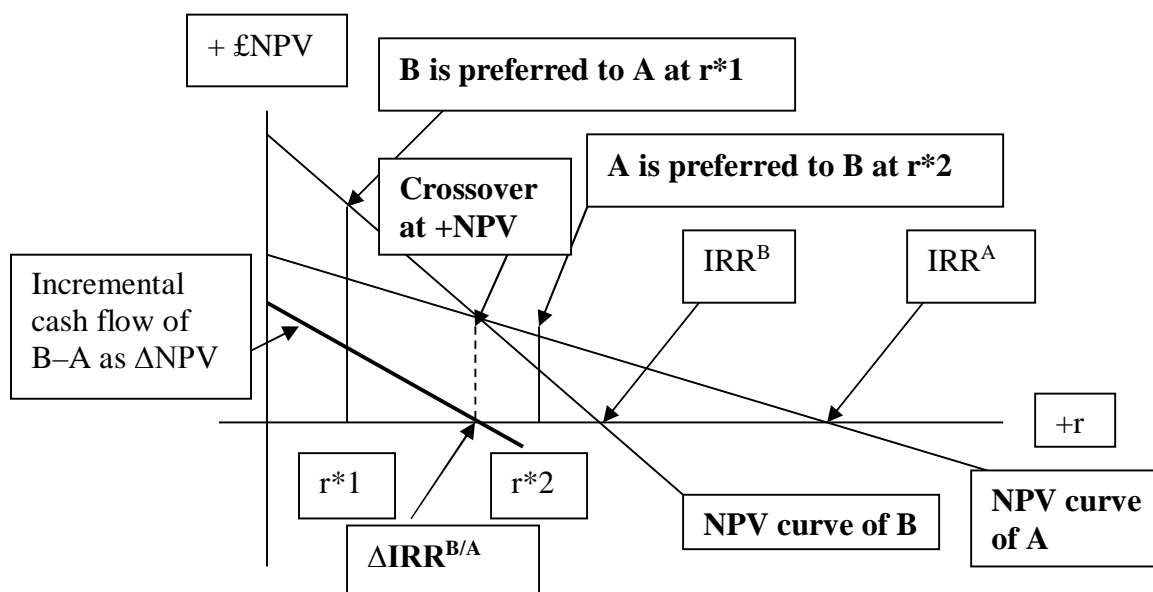
The approach taken, given two ME projects A and B, is:

- * choose the project (say project A) with the highest IRR
- * calculate the 'incremental' cash flow of project B – A
- * calculate the incremental IRR for the incremental cash flow
- * if the $\Delta IRR > r^*$, then choose project B rather than A
- * if the $\Delta IRR < r^*$, then project A is the right choice

The purpose of this is clear from the first chart which shows how the choice of ME project can vary depending on r^* if the NPV curves of the two projects cross-over at some point, whereas the IRR remains the same for both. In this case the IRR of A is the highest but clearly B should be chosen if the cost of capital is r^*1 , but A is the right choice if the cost of capital is r^*2 .

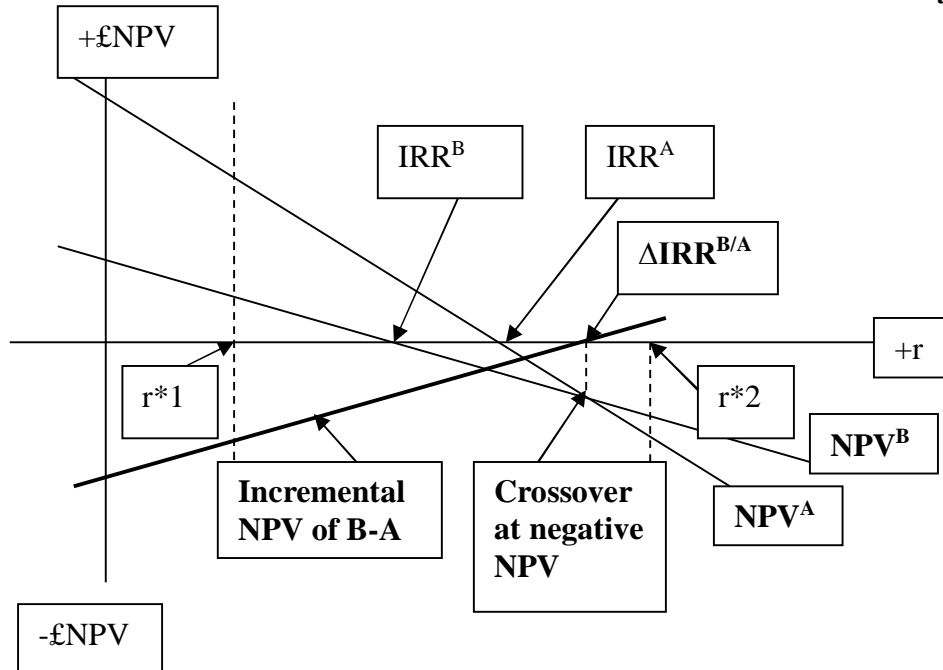
Having chosen A first, you would then change to the choice of B because the incremental IRR of B over A is greater than r^*1 . This can be a tricky approach, however, because the consistency rule based on incremental IRR has to be reversed if the NPV curves of the two projects cross-over at a negative value of NPV. If this is the case, then the incremental NPV curve will be 'upward' sloping. The second chart shows such a position, and once again project A has the highest IRR.

The Incremental IRR consistency test for ME/NCR



In this case, if we choose the project with the highest IRR, project A, then we would want to remain with that choice if the incremental IRR of B-A was greater than r^* , rather than shift to the other project which was the case in the previous example where the projects' NPV curves cross-over at a positive NPV. Whether you use the 'standard' or the 'reverse' rule as the consistency test depends on whether the incremental IRR is greater or less than the two IRRs of the ME projects.

The standard and 'reverse' rule for the Incremental IRR consistency test



The full sequence is therefore:

*Choose the project with the highest IRR (in this case A)

*Do a consistency test based on:

standard rule

if $\Delta IRR^{B/A} < IRR^A$ and IRR^B

Choose B if $IRR^{B/A} > r^*$

Stay with A if $IRR^{B/A} < r^*$

reverse rule

if $\Delta IRR^{B/A} > IRR^A$ and IRR^B

Stay with A if $IRR^{B/A} > r^*$

Choose B if $IRR^{B/A} < r^*$

Separating r^* from the 'z' factor to simplify calculations of TIRR

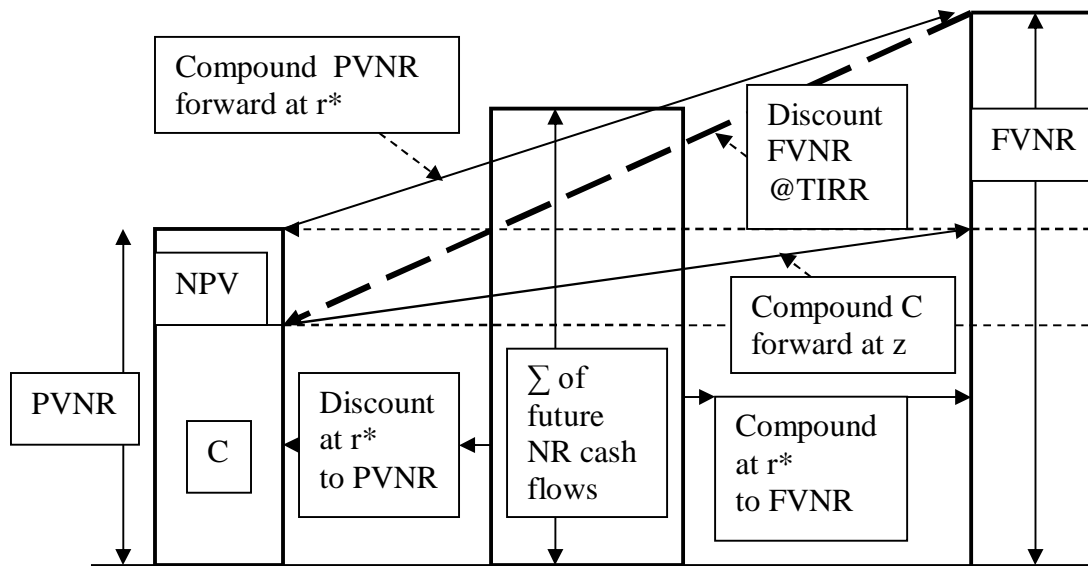
Compounding forward each cash flow element of a cash flow series is the key part in calculating a TIRR (or TIRRC and STIRRC), in order to get its future value at the terminal year for the project or a 'common' terminal year for all projects. This might be seen as tedious computationally, but the process can be simplified because all the necessary information is contained within the NPV and the cost of capital r^* , information you would already have from economic analysis. The approach is to separate the TIRR into two parts, the cost of capital r^* and the 'z' factor, being the annualized rate for the NPV over the life of the project to the terminal year. These parts can be seen as follows, for a capital investment C, with a series of net receipts resulting from it:

$$\text{Given } C(1 + \text{TIRR})^t = \text{FVNR}$$

$$\text{then } C\{(1 + r^*)(1 + z)\}^t = \text{FVNR}$$

Given the NPV has been calculated by discounting the future cash flows at the cost of capital r^* , there is sufficient information to calculate z. The chart shows the relationships, and the formula for z follows.

Showing $(1 + \text{TIRR}) = (1 + r^*)(1 + z)$



The $\text{FVNR} = (\text{NPV} + C)(1 + r^*)^t$ and we want to have a two stage process. To solve for z we set $(C + \text{NPV}) \cdot 1/(1 + z)^t = C$ and therefore $z = \sqrt[t]{((C + \text{NPV})/C)} - 1$. Having solved for z , the $\text{TIRR} = (1 + z)(1 + r^*) - 1$. This makes the calculation of TIRR or TIRRC quite straightforward. To calculate STIRR , we have the capital cost of each ME project plus its residual capital cost to make up the investment sum equal to the capital cost of the most expensive ME project. Its NPV can then be added so that we have, for each project i , $C_i + C_r = C_h$, the investment cost of the most expensive ME project. The STIRR will therefore have a standardised $z = z_s$ and this is solved by setting $C_h = (C_i + C_r + \text{NPV}_i)/(1/(1 + z_s)^t)$, so that $z_s = \sqrt[t]{((C_h + \text{NPV}_i)/C_h)} - 1$.

It is now time to consider the last pane, which concerns how to rank mutually exclusive projects subject to capital rationing, and how to integrate that with independent projects subject to capital rationing.

2. Decision rules using incremental analysis for ME/CR projects

The last pane links the two decision rules, the one for ranking projects subject to capital rationing, and the one for choosing mutually exclusive projects. Incremental analysis links the two by starting with the highest NPV/C and then moving through to the ME project with the highest NPV. With no capital rationing the project with the highest NPV would be chosen. The process of incremental analysis will be looked at first, assuming that all projects are 'strictly time dependent' (STD), that is, need to be built now, or not at all, and to show how a ranking list of independent and mutually exclusive projects subject to capital rationing can be combined into one. It will then be followed by projects which are not strictly time dependent, where the 'opportunity cost of deferral' (OCD) has to be taken into account.

The first step is to list the mutually exclusive projects from the one with the lowest capital costs to the highest, and show their respective NPV and NPV/C calculations. The ME project with the highest NPV/C and the one with the highest NPV can then be identified. These two provide the range of projects which are to be considered in the list. The project with the highest NPV/C is termed the 'base' project and the incremental NPV over the incremental C from that base project can be calculated for each of the remaining mutually exclusive projects. The highest

$\Delta NPV/\Delta C$ is chosen as the new 'base' project and so on until the project with the highest NPV is chosen. Given a set of six mutually exclusive projects, where M2 has the highest NPV/C and M5 has the highest NPV we would have:

MEProject	Cost	NPV	NPV/C	$\Delta NPV/\Delta C$	$\Delta NPV/\Delta C$
M1	x	x	x	n/a	
M2	x	x	xh	'base' project	
M3	x	x	x	M3/M2	
M4	x	x	x	M4/M2h is the new base project	
M5	x	xh	x		M5/M4h is the final project
M6	x	x	x	n/a	n/a

In this case $M2 + \Delta M4/M2 + \Delta M5/M4 = M5$. The list of independent projects (P) ranked according to NPV/C can now be combined with the list of mutually exclusive projects using NPV/C and $\Delta NPV/\Delta C$ to give the single ranking list for comparing with the capital funds available. This could be as follows, showing the accumulated capital cost and three capital constraints:

Project ranking	capital cost	accumulated	capital constraint
P4	50	50	
P7	30	80	
M2	70	150	150
P2	15	165	
$\Delta M4/M2$	20	185	185
P8	40	225	
$\Delta M5/M4$	25	250	NCR

In this case the project choice for each capital constraint would be:

Capital constraint	Project choice
150	P4, P7, M2
185	P4, P7, P2, M4
NCR	P4, P7, P2, P8, M5

Non time dependency and the opportunity cost of deferral (OCD)

The purpose of ranking projects subject to capital rationing is to determine the opportunity cost of using a scarce resource, in this case capital funds. NPV is a measure of opportunity cost, but only where the project is strictly time dependent (if you do not undertake the project now, its NPV is lost). However, there are many situations in which a project can be 'deferred', and assuming that the resulting cash flow would be the same, irrespective of whether you started now or in a year's time, in this case the NPV is not lost, and the opportunity cost of not undertaking the project now is simply the time cost of deferral. The time cost of deferring for one year is simply the difference between the NPV starting now and the NPV starting next year, discounted at the cost of capital r^* , that is, $OCD = NPV(1 - df_1)$. This would be a small figure compared to the NPV, and is clearly important in choosing projects in a year subject to capital rationing, where it is more important to focus on the strictly time dependent projects.

However, for many projects that can be deferred, it may still be considered risky, and such risk can be incorporated into the calculation of opportunity cost with a 'deferral risk multiplier' (m). The multiplier can be said to range between nought and $1 + r^*$. Where the multiplier exceeded $1 + r^*$, then this is saying that the NPV calculated for a project increases as it is deferred, so that

it would be postponed anyway (and reflects the fact that there is little point in spending money now if the benefits are not expected to accrue until some time later). Where $m = 1$, then the project is strictly non-time dependent (SNTD), and where $m = 0$, then it is strictly time dependent. Where m falls somewhere in between, it is a measure of the risk of deferral. In the window on investment appraisal therefore, the NPV is best reflected more generally as opportunity cost where $OCD = NPV_0(1 - m.df_1)$ for a two year time horizon. This results from comparing the NPV from starting now (0) and the NPV from starting in a years time (1), that is $NPV_1 = m.NPV_0$, so that $OCD = NPV_0 - m.NPV_0.df_1$.

Preclusion and mutually exclusive projects

The final step before we can complete in full the panes of the window on investment appraisal is to take account of the 'preclusion' opportunity cost for deferrable mutually exclusive projects subject to capital rationing. It is not enough just to have converted NPV into OCD as above for the ranking list calculations, as OCD has to be adjusted for the fact that if a mutually exclusive project which does not have the highest NPV is to be chosen out of the ranking list, then account must be taken of the fact that it 'precludes' the possibility of undertaking a higher NPV value ME project next year. The conclusion is that in calculating the opportunity cost for ranking purposes, the 'preclusion' cost of excluding other incremental ME projects must be included.

The introduction of a risk deferral multiplier has also to take account of the fact that if the m is not a common constant for all the mutually exclusive projects in the set to be chosen from, then the ranking to find the highest NPV if the project is started now might not be the same as the one with the highest NPV if started next year. The OCD for each project, including the preclusion element, has then to be based on the project with the highest NPV next year. The formula can be illustrated using four mutually exclusive projects (1 to 4) with project 1 having the highest NPV/C (termed R), project 4 having the highest NPV if started now, and project 3 having the highest NPV if started next year, taking account of the risk deferral multipliers m , each being shown in bold for comparison. The data table would be:

MEP	capital cost	NPV starting year 0	NPV/C	NPV starting year 1
1	C1	NPV1	R1	NPV1. <i>m</i> 1
2	C2	NPV2	R2	NPV2. <i>m</i> 2
3	C3	NPV3	R3	NPV3.<i>m</i>3
4	C4	NPV4	R4	NPV4. <i>m</i> 4

and the formula for the OCD of each ME project would be its NPV_0 less the time element of deferral and the preclusion element, given $NPV_1 = NPV_0.m$ for each project:

$$OCD = NPV_0 - NPV_0.m.df_1 - (\text{highest } NPV_0.m - NPV_0.m).df_1$$

By rearranging this simplifies to:

$$OCD = NPV_0 - (\text{highest } NPV_0.m).df_1$$

This formula allows you to calculate the 'residual' value of doing something now rather than a year later, that is, the project's opportunity cost of deferral. This can be illustrated with a simple worked example for two mutually exclusive projects (1 and 2), given a discount rate r^* of 10%, and comparing them with two sets of risk multipliers m , option 'a' being 0.8/0.2, and option 'b' being 0.5/1 respectively:

MEP	C	NPV ₀	NPV ₀ /C	NPV _{1a}	NPV _{1b}
1	80	20	.25	(20)0.8 = 16	(20)0.5 = 10
2	100	22	.22	(22)0.2 = 4.4	(22)1 = 22

In this case there is a reversal in terms of the mutually exclusive project with the highest NPV at the start of next year, depending on the risk factor applying to each. The OCD for each, depending on whether it is option a or b, is:

MEP	OCD for option a	OCD for option b
1	$20 - 16(0.909) = 5.456$	$20 - 22(0.909) = 0.002$
2	$22 - 16(0.909) = \mathbf{7.546}$	$22 - 22(0.909) = \mathbf{2.002}$

and so the ranking on OCD/C would be:

MEP	OCD/C for option a	OCD/C for option b
1	$5.456/80 = 0.0682$	$0.002/80 = 0.000025$
2	$7.546/100 = \mathbf{0.07546}$	$2.002/100 = \mathbf{0.2002}$

and in this case, although mutually exclusive project 1 had the highest NPV/C for a ranking list, in terms of OCD, on both options, project 2 has the highest OCD and OCD/C. In this case there would be no need for incremental analysis on the base of project 1, which had the highest NPV/C. If it was calculated, the $\Delta\text{NPV}/\Delta\text{C}$ on the base of project 1 would have been negative.

It is interesting to note that whilst the OCD is superior to using NPV when ranking independent and mutually exclusive projects subject to capital rationing (because it pushes deferrable projects further down the ranking list relative to projects that are strictly time dependent), the calculation of incremental OCD and incremental NPV for mutually exclusive projects remains the same. This is because the OCD formula above reduces the NPV of each ME project by a common amount (CA):

Given $\text{OCD} = \text{NPV} - \text{CA}$ then the incremental OCD for two ME projects, 1 and 2, is
 $(\text{NPV}_2 - \text{CA}) - (\text{NPV}_1 - \text{CA}) = \Delta\text{OCD}$, which by rearranging is
 $\text{NPV}_2 - \text{NPV}_1 = \Delta\text{OCD}$, which also $= \Delta\text{NPV}$.

To illustrate this, take two mutually exclusive projects 1 and 2, which have NPVs of 40 and 100 respectively, and capital costs of 30 and 80 respectively. In this case the NPV of project 2 is the highest, but project 1 has the highest NPV/C ($40/30 = 1.333$). The incremental NPV/C of 2 over 1 is $60/50 = 1.2$. The NPV/C of project 2 is $100/80 = 1.25$ and this number is the 'weighted' average of the NPV/C for project 1 and the incremental NPV/C of project 2 over project 1 (ie, 1.25 compared to 1.333 and 1.2). If project 2 had the highest NPV, say 70, then the OCD for each project would be:

For project 1 $40 - 70 = -30 = \text{OCD}_1$
For project 2 $100 - 70 = 30 = \text{OCD}_2$

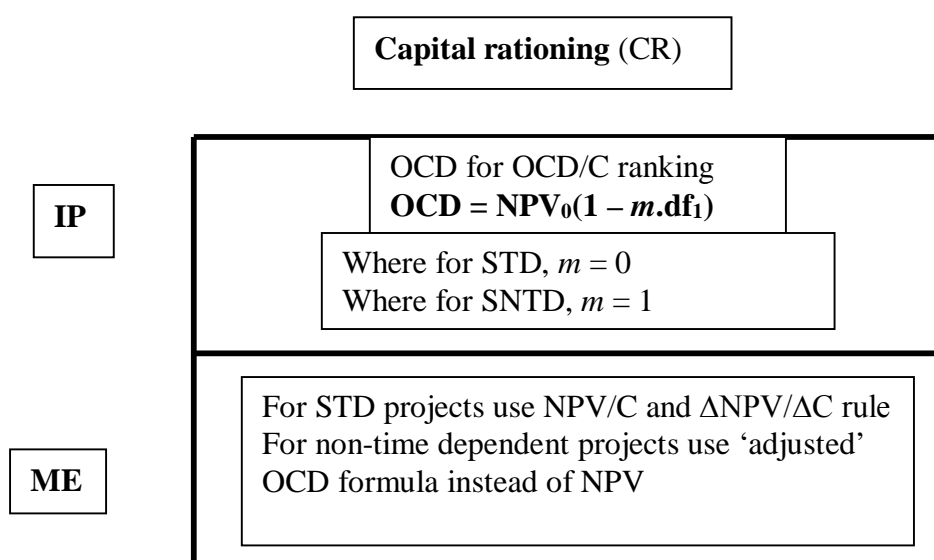
In this case project 2 would have the highest OCD and the highest OCD/C, and clearly it would have been incorrect to have selected project 1 to be started now just because it had a higher NPV/C and there were insufficient capital resources to allow project 2 to be started. The incremental OCD calculation of project 2 over 1 is unnecessary in this case, because project 1 has already been removed from the ranking list, but if it was done it would be $(\text{OCD}_2 - \text{OCD}_1)/(\text{C}_2 - \text{C}_1)$, which is $((30 - (-30))/(80 - 30))$, which is $60/50 = 1.2$, the same figure as the incremental NPV/C above.

With the possibility of deferring projects subject to capital rationing, it is clear that NPV/C is not the appropriate ranking method for projects subject to capital constraint, except where all the projects are strictly time dependent (do now or not at all). The question arises, however, as to

what happens if there is also a capital constraint applying to next year, and possibly years thereafter, and with additional projects listed to be started in year 2, 3 and so on. In this case, one approach would be to do a 'step-wise' approach, choosing the projects to be started in year 1 based on the OCD analysis of each project in year 1. Having chosen the projects for year 1 which are best started then, and in total satisfy the capital constraint for the year, move on to year 2 and consider the projects listed to be started in year 2 plus the deferred projects from year 1. Having completed year 2, move on to year 3, and so on. If projects are deferred more than one year then account has to be taken in calculating OCD of the risk deferral multiplier m appropriate for the year in question. The selection of projects for each year may therefore include projects listed to start in that year plus deferred projects of one or more years.

The two window panes on investment appraisal related to capital rationing in the current year can now be as shown in the chart, where the more general ranking measure of opportunity cost of deferral (OCD) has replaced NPV:

The window panes for capital rationing using OCD



Whilst OCD and modified IRR analysis has allowed us to provide a full set of panes for the window on investment appraisal, and to provide a more general presentation than is typically found in many textbooks, in capital rationing we have noted that there might be more than one constraint, rather than just capital rationing in this year, with no capital constraints set for subsequent years. If there were, the step-wise procedure shown can address the problem of 'multiple' capital constraints, but of course there are many other constraints that might apply in the production and budgeting process, such as scarce labour and machine time, which have to be taken account of in looking for the best management decision which increases revenues and reduces cost in such a way that NPV is maximised, subject to meeting the various constraints. This leads us to the subject of optimising with multiple constraints using linear programming.

3. Optimising with Multiple Constraints and 'Dual' Solutions

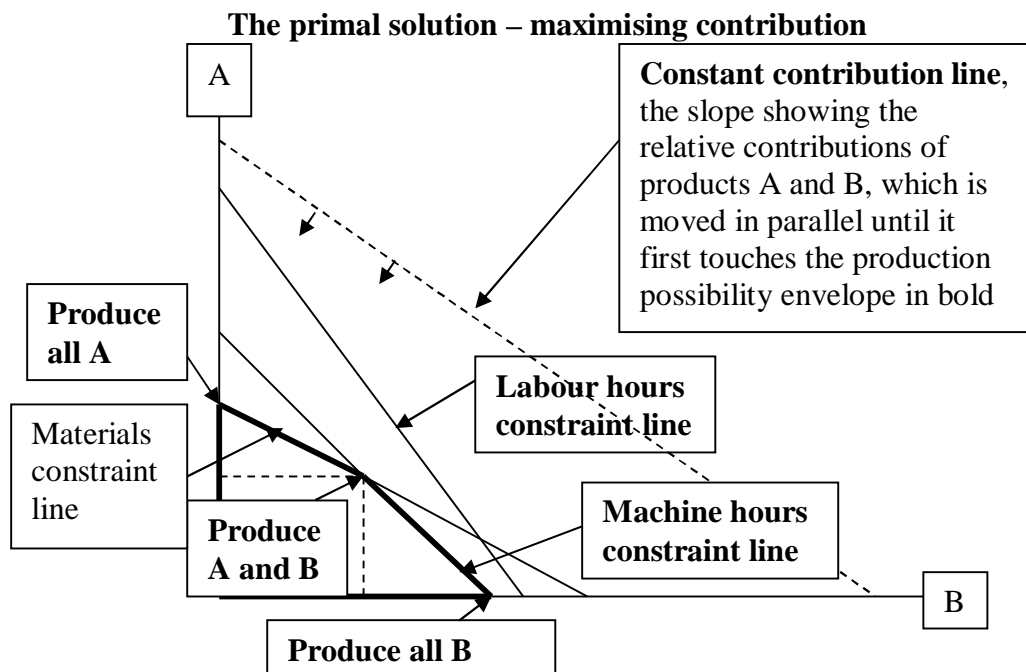
A management decision might be concerned with how many of a range of different products to produce and sell, subject to a number of constraints on how many of each product can be produced. The contribution per unit sold might differ for each product, and each product might use different amounts of the constrained resources which have to be used in producing the

products. Linear programming is a process to identify which combination of products 'maximises' the contribution (C), subject to the constraints. It can also be used to identify which combination of inputs can 'minimise' the cost of production which is subject to certain constraints, such as the chemical mix.

The process of linear programming can best be illustrated using a choice of two products (A and B) subject to three constraints (labour hours, machine hours, and material resources – L, M and R). The essential element is to establish the 'objective' function (in this case to maximise contribution), and to list the set of constraints (for each, the amount used to produce one unit of each product and the total available for use). The assumption is made that each product is perfectly divisible (eg, if the primary unit for butter is 1lb, butter can still be sold in half pound units). Other forms of programming are available where divisibility is not physically possible, such as 'integer' programming. In the simple case we are using, the linear programme would be written as follows to solve for A and B:

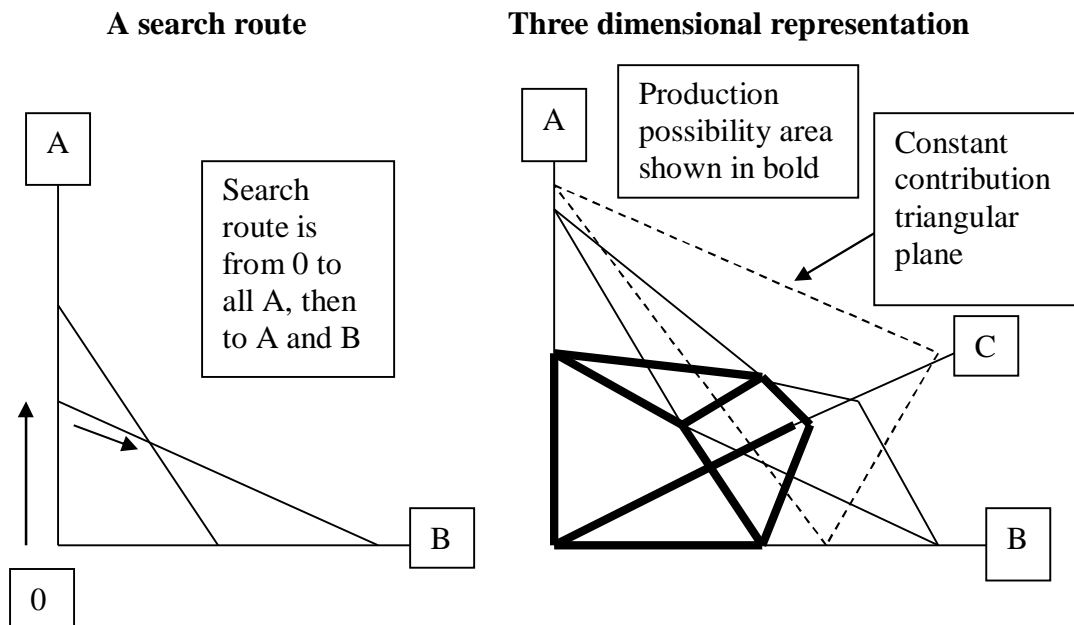
Maximise (contribution)	$C^A.A + C^B.B$
Subject to (constraints)	
Labour hours	$L^A.A + L^B.B \leq L^T$, the total labour hours available
Machine hours	$M^A.A + M^B.B \leq M^T$, the total machine hours available
Resource use (materials)	$R^A.A + R^B.B \leq R^T$, the total of the resource available

which means that we are looking to choose a combination of A and B (which could be all A, all B or a mix of A and B) which maximises the contribution while at the same time satisfying the constraints (ie, less than or equal to the total available in each case). In this case, as we would not want negative values for products A and B, there would have to be non-negative constraints too in order to properly delimit the feasible production area (ie, $A \geq 0$ and $B \geq 0$). An illustrative diagram shows how the information above is used in the 'search' technique of linear programming. The first step is to plot the 'feasible' production area, given the constraints, and then to use the relative contributions of the two products to identify which production mix is optimal.



The constraint lines show the maximum combinations of A and B which can be produced, and so the production feasibility area lies in the area between the points all A, through a mix of A and B, to all B. In this case, the materials and machine hours constraints are the 'binding' constraints, and the labour hours constraint is 'non-binding' because there are always extra labour hours available at any point on the feasible production area, which is limited purely by the binding constraints. Which of those three points is the best depends on the slope of the constant contribution line, the aim being to move the constant contribution line as far away from the origin as possible while still touching the production possibility area. In the illustration, clearly the constant contribution line is well beyond the production possibility area. Moving it back towards the origin in parallel to the slope of the constant contribution line, it would first touch the production possibility area at point A and B (a production decision to produce a mix of A and B, in this case because the slope of the constant contribution line falls between the slopes of the materials and machine hours constraint lines). If not, the optimal decision would have been to produce all A or all B.

The simple case of two products and three constraints can usefully be drawn to illustrate the relationships between the constraint lines and the relative contributions in searching for the optimal production decision. In practice the number of products included in the production mix is usually equal to the number of binding constraints. The effect of three products and two constraints is usefully shown, and the illustration is also useful to showing the search process which linear programming involves. In the three product illustration, which combination (all A, all B, all C or a mix of A/B or A/C) will prove best depends on the orientation of the triangular constant contribution line for A, B and C. It soon becomes clear that as the number of products and constraints increases, it is best to leave computers to calculate the optimal point. An example of the 'simplex' search method is shown in the Appendix.



Minimising cost as the objective function

The same approach is used when the objective is to minimise a cost, subject to constraints. An example might be choosing a mix of feed for animals, given certain nutritional requirements, or designing a cost-effective, but nutritional school meal. The only difference is that the objective function is minimised, and the constraints must be 'greater or equal to' the minimum constraint

allowed. The following is to choose a combination of apples (A) and/or bananas (B) which is the cheapest but meets the minimum nutritional requirements of calories and vitamins.

Minimise cost (choice of fruit)

$$P^A.A + P^B.B$$

Subject to constraints

Vitamins (V)

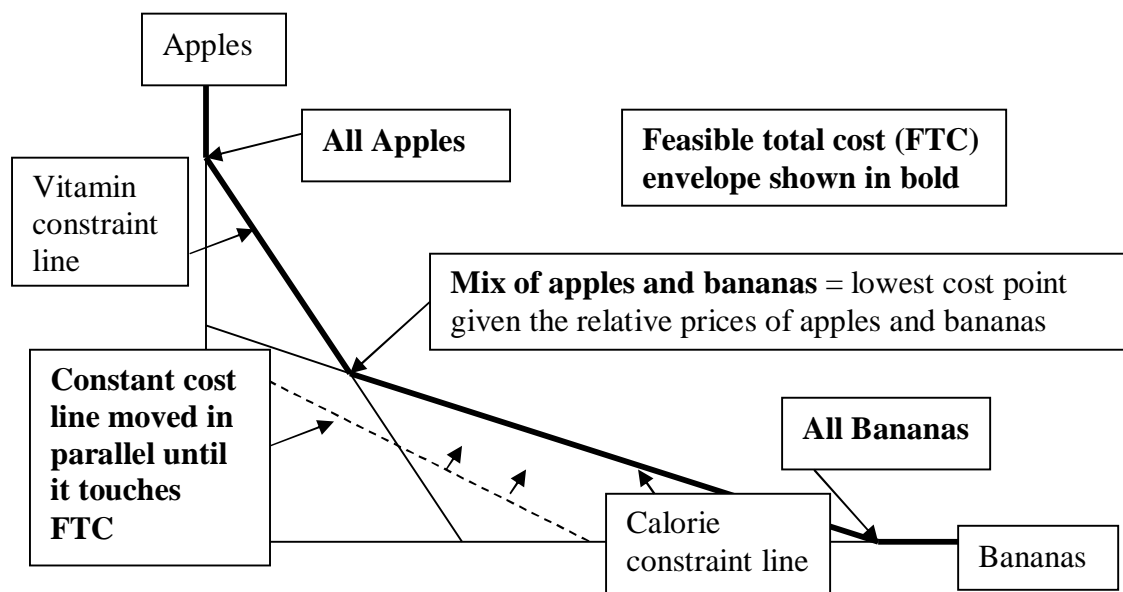
$$V^A.A + V^B.B \geq V^T, \text{ the minimum total vitamins}$$

Calories (C)

$$C^A.A + C^B.B \geq C^T, \text{ the minimum total calories}$$

In illustrating this minimising example, the 'feasible cost' area is now an 'envelope' above the minimum constraint lines, and the objective is to move the constant cost line as close to the origin as possible, while still touching the feasible cost area which satisfies the constraints. In this case, because of the slope of the constant cost line, the minimum cost choice of fruit is a mix of bananas and apples.

The primal solution – cost minimisation



The final point to make before looking at the dual solutions is that linear programming can turn the inequalities of the constraints into 'equalities' by introducing a 'slack' variable into each constraint. A binding constraint would have a zero slack variable but not a non-binding constraint. In the simple maximising example above, the slack variable for labour hours would be the number of labour hours still available once the optimal production mix had been chosen. In these simple examples, the binding constraints are equalities, since all the scarce resource is used up, so the optimal solution can be solved from a simultaneous equation.

Dual solutions and opportunity costs

The linear programming approach is useful to either maximise contribution earned, or minimise the cost of supply, given a set of constraints. An alternative question naturally arises, which is, given the constraints, what is the 'opportunity' cost of each of the constraints, which would be helpful in deciding whether to purchase more of the constraining resources. Binding constraints have opportunity costs, and non-binding constraints have no opportunity costs because they are surplus to requirements. One approach to this is to solve for the opportunity costs by using the primal method as above and comparing the results where the first solution is based on the constrained resources available, and the second solution is calculated with one of the constraints having one extra unit. The difference between the two solutions (say, increase in the total contribution from having one extra unit of a constrained resource) is the 'opportunity' cost of

not having that extra unit of constrained resource. If the opportunity cost exceeds the cost of acquiring a further unit then it is worth purchasing more, and so on until the opportunity cost equals the cost of acquiring more of the constraining resource.

The 'dual' solution to a linear programme is very useful because it simply reflects the reordering of the terms in the primal solution. We can see this using the two products (A and B) and three constraining resources (L, M and R) example above. Whereas we were solving for A and B to achieve the highest total contribution, in the dual case we are solving for the opportunity costs of L, M and R (that is l, m and r) which, when multiplied by their respective constrained resource totals, would equal the maximum contribution which resulted from the primal solution. The relationship is directly noticeable from setting the primal and dual solutions side by side, given maximise becomes minimise, less than or equal to becomes greater than or equal to, and columns become rows and vice versa.

Primal (solving for A and B)				Dual (solving for l, m and r)			
Max	$C^A.A + C^B.B$			Min	$l.L^T + m.M^T + r.R^T$		
St	L	$L^A.A + L^B.B \leq$	L^T	St	A	$l.L^A + m.M^A + r.R^A \geq$	C^A
	M	$M^A.A + M^B.B \leq$	M^T		B	$l.L^B + m.M^B + r.R^B \geq$	C^B
	R	$R^A.A + R^B.B \leq$	R^T				

For the dual solution we have three terms to solve for, with two constraints, compared with the primal which had two terms to solve for and three constraints. The dual solution shows that when the opportunity costs of L, M and R are solved, then the sum of the opportunity costs for a unit of each product A or B must be equal to or greater than the contribution available from selling a unit of A or B. Since it would be uneconomic to produce at the margin where the total opportunity cost of its use of constrained resources was greater than the contribution it could make, the solution of the dual linear programme will necessarily relate to the solution of the primal linear programme as follows: for products included in the production plan solution of the primal programme, the product constraint lines for those products in the dual solution will be equalities; where the opportunity costs are greater than the contribution, these products will have been excluded from the maximising production mix of the primal programme.

The dual solution is therefore consistent with the maximising solution, and provides necessary economic information in terms of the opportunity costs of scarce resources. It is because the dual solution is an exact reflection of the primal solution that economists refer to the results as 'shadow' prices. In practical terms it is necessary to recognise that the calculations are 'marginal' opportunity costs around the area of the optimal production mix solution. If continuing further units of a scarce resource are added, it may be that at some point it ceases to be a binding constraint, and its opportunity cost falls to zero. The total contribution from that point onwards has to be distributed amongst the remaining binding constraints, and their opportunity costs will change.

Illustrative worked example

Paganini Ltd is considering how much of four products (M,N,R and S) to produce in the next period, given it has labour hours' constraints for the three grades of labour it uses (A,B and C). The limits are 9000 hours for Grade A, 14500 hours for Grade B, and 12000 hours for Grade C. Each grade of labour is paid £6 per hour, but they are not substitutable given the different skills required. There are no constraints on the amount of each product that can be sold. The management is therefore interested to know what the cost of changing their labour hours' constraints would be, and wants to be sure that the maximum contribution they can achieve

covers the fixed common costs which need to be recovered of £142,000. The management accounting information they have is:

Product	M	N	R	S
Selling price per unit	220	212	388	344
Cost of materials per unit	68	100	152	88
Variable O/H costs per unit	24	28	20	24
Labour hours per unit – Grade A	10	6		
Grade B	-	-	10	20
Grade C	-	-	12	6

From this information, the management accountant works out the contribution per unit of M, N, R and S, and the contributions per scarce labour hour used.

	Contribution	Contribution per labour hour used
Product M	$220 - 68 - 24 - 10 \times 6 = \mathbf{68}$	Grade A: $68 \div 10 = 6.8$
Product N	$212 - 100 - 28 - 6 \times 6 = \mathbf{48}$	Grade A: $48 \div 6 = \mathbf{8}$
Product R	$388 - 152 - 20 - (10+12) \times 6 = \mathbf{84}$	Grade B: $84 \div 10 = \mathbf{8.4}$
		Grade C: $84 \div 12 = 7$
Product S	$344 - 88 - 24 - (20+6) \times 6 = \mathbf{76}$	Grade B: $76 \div 20 = 3.8$
		Grade C: $76 \div 6 = \mathbf{12.667}$

The contribution table shows that product N is preferred to product M in terms of its use of Grade A labour, but for products R and S there is a ‘reversal’ of ranking with product R being preferred to product S in terms of its use of Grade B labour, but product S is preferred to product R in terms of its use of Grade C labour. This means it would be best to produce as much of product N as possible and a combination of products R and S. The linear programme to determine the combination of products R and S would be:

$$\begin{aligned} &\textbf{Maximise } 84R + 76S \\ &\textbf{Subject to: } \text{Grade B labour } 10R + 20S \leq 14500 \\ &\quad \text{Grade C labour } 12R + 6S \leq 12000 \end{aligned}$$

and because of the reversal in ranking we know that both constraints will be binding, therefore the inequalities in this case can be made equalities. The quantities of R and S can therefore be solved from a simultaneous equation. Solving for this, $R = 850$ units and $S = 300$ units to be produced.

The financial outcome from producing this many units of N, R and S would be:

	quantity	contribution	surplus
Product N: $9000 \text{ hours} \div 6 \text{ hours per unit} =$	1500	$x \quad 48$	$= \quad 72,000$
Product R:	850	$x \quad 84$	$= \quad 71,400$
Product S:	300	$x \quad 76$	$= \quad \underline{22,800}$
		Total contribution =	166,200
		less fixed common costs	<u>142,000</u>
		<i>equals</i> total surplus	24,200

and the fixed common costs have been recovered.

The question of how much the management would be willing to pay to release the constraints, given labour costs £6 per hour can be answered as follows:

For Product N the contribution per hour is $48 \div 6 = 8$, therefore they would be willing to pay up to £8 per hour

For products R and S we need to calculate the ‘internal’ opportunity cost of grade B labour (b) and grade C labour (c) from the ‘dual’ linear programme, which again can have equalities because both products R and S are in the optimal production plan:

$$\begin{array}{ll} \text{Minimise} & 14500b + 12000c \\ \text{Subject to: Product R} & 10b + 12c \geq 84 \\ \text{Product S} & 20b + 6c \geq 76 \end{array}$$

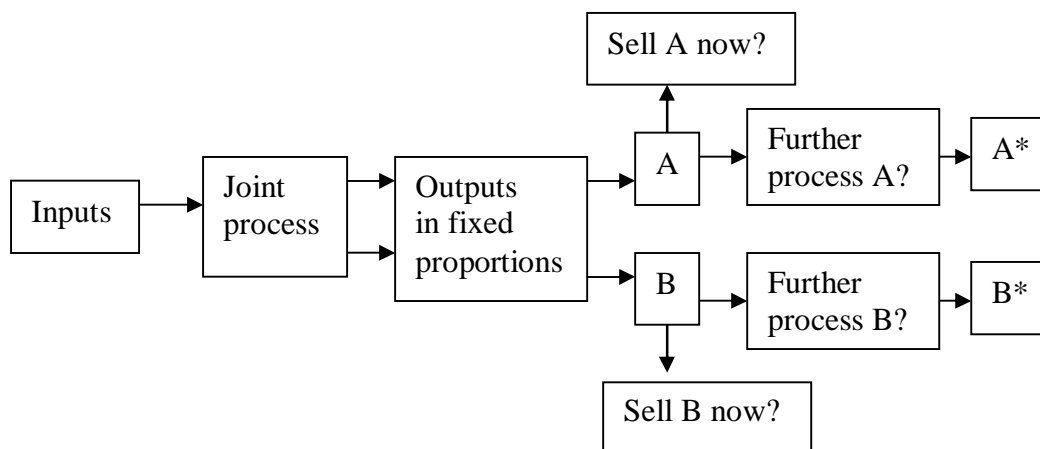
and the solution to this for b and c, given ‘equalities’, means that the opportunity cost of b is 2.266667 and for c is 5.111111. These are the extra amounts over £6 that management would be willing to pay for Grade B and Grade C labour per hour. The calculation of ‘internal’ opportunity cost means that the contribution per unit of products R and S is fully allocated to the constraining resources, grade B and C labour. For this reason $14,500 \times 2.266667 + 12,000 \times 5.111111 = £94,200$ the maximum contribution from selling products B and C.

4. Economic allocation of joint costs

Joint costs are a useful point at which to combine traditional accounting methods with economic analysis. Joint costs arise when there are multiple product outputs from a set of inputs into a production process where the outputs remain in fixed proportions to the inputs. An example would be the input of crude oil into a refining process which results in a range of fuels being produced. Accountants have typically been concerned with how the costs of the joint process should be allocated to each of the outputs, and economists with deciding on the optimal level of production from the joint process.

The management accounting issues arising from joint costs are best illustrated in the following diagram, which also illustrates that decisions can be made to ‘further process’ one or more of the joint outputs (A and B) to yield more valuable products (A* and B*). Whether to further process will depend on whether the cost of further processing is greater than or less than the difference in value between A* and A and B* and B. The costs which should be attributed to each product in this case are clear (eg, further processing costs on A are attributed to A*). The problem for accountants is how to allocate the costs of the joint process to A and B. Economists are unconcerned because for them the only problem is to ensure that the right decisions on further processing are made, and given the products to be produced, the total revenue can be compared with the total cost of both the joint process and the further processing.

Joint Process Production and Further Processing



Three choices are usually provided by accountants, and the most favoured is to allocate joint process costs to each output in proportion to their realisable value (£s). This has a particular

advantage over the other method, which is to allocate in proportion to the physical output of each (kg, litres). The problem with this method is that the differences in the distribution of realisable values and physical values means some products could become 'loss-making', the allocated costs exceeding their realisable value. Given the joint process overall is worth undertaking, such accounted losses are meaningless, given all the outputs are supplied in given fixed proportions. The remaining method is an extension of the realisable value method, whereby the costs of the joint process are allocated to the outputs in proportion to the realisable value of the final outputs produced, that is, including any further processing (so, for example, rather than allocating the joint costs in proportion to the realisable value of A and B, if there has been further processing of A, then the costs are allocated in proportion to the realisable value of A* and B). This method can be useful if the joint process is not worthwhile unless some of the outputs are further processed and increase the overall value of the output.

Economic allocation

Given that the different product outputs of a joint process might be subject to very different demand, and there is also the need to use that information to decide on the optimal production level and pricing, the opportunity arises to combine both economic appraisal and cost allocation in one. In this case the allocation of joint costs will be proportional to the marginal revenues of each of the output products at the optimal level of production inputs. The first step is to find the production point where the marginal cost of the joint process is equal to the sum of the marginal revenues of the outputs. The optimising rule can be found because the outputs are in fixed proportions and therefore the demand curves can be expressed in terms of inputs. Given a unit of input I, then each product output has an input/output relationship 'phi' (Φ), so that we can substitute Q in each of the demand curves by its appropriate Φ . For example, if there are two units of product A for each unit of input I, then $Q^A = 2I$ because $O^A/I = \Phi_A$.

The optimising production level will be where the marginal cost of the joint process for a unit of input I (MC^{jp}_I) equals the sum of the marginal revenues of the outputs resulting from an input of I. Given straight line demand curves for each product the the standard marginal revenue in terms of small changes of Q would be $MR = a - 2b(\Phi I)$. However, we are concerned with small changes of input I, so the relationships would be as follows:

$$\text{Average revenue (price)} = a - b \cdot \Phi \cdot I$$

$$\text{Total revenue (P.Q)} = a \cdot \Phi \cdot I - b \cdot \Phi^2 \cdot I^2$$

$$\text{Marginal revenue (dTR/dI)} = a \cdot \Phi - 2b \cdot \Phi^2 \cdot I = \Phi(a - 2b \cdot \Phi \cdot I)$$

and it can be seen that the traditional marginal revenue for each output has to be multiplied by its input/output relationship Φ . The optimising rule is therefore set:

$$MC^{jp}_I = \Phi_A(MR^A_I) + \Phi_B(MR^B_I) + \dots\dots\dots$$

Primary products and By-products

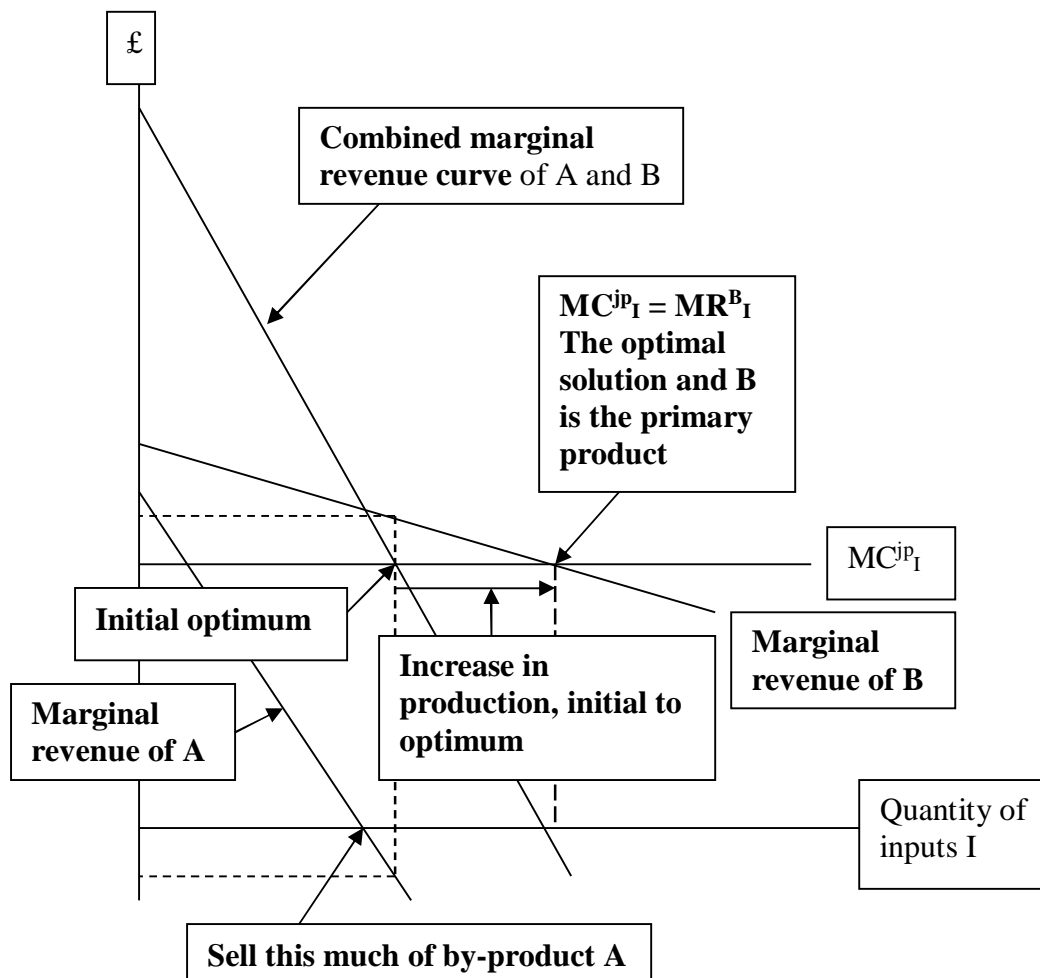
This formula raises an interesting question, which requires the separation of the outputs into primary products, that is, those products which are necessary to determine the optimal level of production, and the remaining products which do not, being By-products. This arises because the above formula could be solved with some marginal revenues which are positive and some which are negative. Those which are negative become by-products and are eliminated from the optimising equation. In essence the process goes through a second round and the optimal production level increases. Once this has been solved, it is necessary to check whether the increased production has turned products with a positive marginal revenue in the first round into products with a negative marginal revenue in the second round. If so, those are then eliminated

and the equation solved again as a third round. The process is complete when the optimal solution has products all of which have positive marginal revenues.

As the joint process will give a fixed proportion of by-products from the optimising solution based on primary products alone, the question arises about how to price the by-products and allocate the joint costs. The answer is to sell by-products up to the point where their marginal revenue is zero (which maximises total revenue), and this can be done where the marginal cost is also zero, that is, no joint costs should be allocated to by-products, the joint cost being allocated to primary products in proportion to their marginal revenues at the optimum point. Surplus by-products are not sold but disposed of appropriately. Economic allocation of joint costs is therefore to be preferred to the accounting method of allocating in proportion to realisable value, which is effectively using price, that is, average revenue rather than marginal revenue. The following diagram based on two outputs A and B for inputs I illustrates the relationship between primary products and by-products, and shows how the initial optimum changes once the by-product is eliminated. The marginal cost of the inputs is assumed to be constant (that is, $AC = MC$).

The diagram shows how the initial optimum is a result of combining an excess of marginal revenue for B with a negative marginal revenue for A. When the by-product A is eliminated, then production increases until the optimal solution where the marginal revenue of B equals the marginal cost of the joint process.

Optimal production, primary products and by-products



Illustrative worked example

John Outback is a sheep farmer who uses his sheep to produce wool for sale and then be sold to butchers for mutton. The size of his herd of sheep is fairly constant, with new lambs being used to replace sheep finally sold for their mutton, but the size can change as demand changes. He has found that the quality of his wool and mutton is improved if he uses a fertilizer on a regular basis and which has the following minimum elements per bag: 5 kilos of general fertilizer (g), 25 grams of vitamins (v), and 12 milligrams of trace elements (t). He has found that the quantities of wool and mutton come in fixed proportions so that on average 1 unit (bag) of fertilizer results in half a unit of wool and 3 units of mutton. The variable costs associated with that (including mixing, shearing etc) come to £10 per outputs from 1 bag of fertilizer in the joint process.

John Outback wants to review his herd size and purchase of fertilizer now that the supplier of fertilizer has introduced two grades of fertilizer; Premium grade (P) for £4.5 a bag and Standard grade (S) for £3 a bag. Each grade has the following elements, which can be compared with the minimum requirements for John Outback:

	general fertilizer	trace elements	vitamins
premium grade	5 kilos	16 mgs	50 grams
standard grade	5 kilos	8 mgs	12.5 grams

The premium grade bag provides enough of each to cover the minimum requirements of John Outback, but is expensive, and provides more than the necessary amount of vitamins and trace elements. By mixing some standard with premium grade fertilizer, it should be possible to reduce the cost but still meet the minimum requirements. The minimizing linear programme to do this would be:

	Minimise	$4.5P + 3.0S$
Subject to:	General fertilizer	$5P + 5S \geq 5$
	Vitamins	$50P + 12.5S \geq 25$
	Trace elements	$16P + 8S \geq 12$

and to solve this it is necessary to ask which are the 'binding' constraints where the inequalities can be turned into equalities. This will depend on the relative amounts of a bag of premium grade which is necessary to provide each of the three elements, and the 'gradients' for transferring from premium to standard grade for each of the elements and the 'constant cost' line. The figures for this are as follows:

	general fertilizer	trace elements	vitamins
relative proportion	$1/1 = 1$	$12/16 = 0.75$	$25/50 = 0.5$
gradients S/P	$1/1 = 1$	$1.5/0.75 = 2$	$2/0.5 = 4$

and the constant cost gradient is $1.5/1 = 1.5$

Given the general fertilizer and the trace elements require a higher proportion of the premium grade bag to meet the minimum requirements than do the vitamins, and because the gradient of the constant cost line falls between the substitution gradients for general fertilizer and trace elements, it will be the general fertilizer and trace elements which will be the binding constraints, not the vitamins. This can be usefully demonstrated visually if the gradients are drawn on graph paper. From the linear programme we can therefore create a simultaneous equation for the binding constraints and solve for P and S as follows:

$$\begin{array}{rcl} 5P + 5S = 5 & \text{can become} & 8P + 8S = 8 \\ 16P + 8S = 12 & \text{less} & \underline{16P + 8S = 12} \\ & \text{equals} & 8P = 4 \end{array} \quad \text{therefore } P = 0.5 \text{ and } S = 0.5$$

The minimum cost of the mixed bag of fertilizer which meets the constraints is therefore $0.5 \times 4.5 + 0.5 \times 3 = \text{£}3.75$

Having decided how to purchase his fertilizer, the next question is to decide the amount of fertilizer (F) to buy each month to meet the monthly demands, given the joint process for wool and mutton. The monthly demand curves are for wool $P^W = 60 - 0.5Q^W$ and mutton $P^M = 20 - 0.2Q^M$, and given the output/input ratios for wool as 0.5 (ie, $Q^W = 0.5F$) and mutton as 3 (ie, $Q^M = 3F$), the marginal revenues in terms of F, the input of fertilizer, can be calculated:

	Wool	Mutton
Average revenue in terms of Q	$60 - 0.5Q^W$	$20 - 0.2Q^M$
Average revenue in terms of F	$60 - 0.25F$	$20 - 0.6F$
Total revenue in terms of F	$60 - 0.25F^2$	$20 - 0.6F^2$
Marginal revenue in terms of F	$60 - 0.5F$	$20 - 1.2F$

Set the sum of the marginal revenues per unit of F equal to the sum of the cost of a unit of F plus the variable costs associated with the joint process outputs: $60 - 0.5F + 20 - 1.2F = 3.75 + 10$. Solving for F would give an input of 38.970588 bags of fertilizer a month, but the marginal revenues of wool and mutton associated with that input would be for wool = + 40.514706 but for mutton it would equal - 36.117646, therefore mutton is a by-product of the joint process. Eliminating the marginal revenue of mutton, and setting $60 - 0.5F = 13.75$, the optimal input of fertilizer in the month would be $F = 92.5$.

From 92.5 inputs of F there would be 46.25 units of wool and 277.5 units of mutton. The wool would be sold at a price of £36.875 per unit of wool (ie, $60 - 0.5 \times 46.25$) and the mutton would be sold at a price per unit where the marginal revenue of mutton equalled zero, that is, £10 and at which point the quantity sold would be 50 units of mutton (ie, $20 - 0.4Q^M = 0$). Given 277.5 units of mutton arose from the joint process, in order to maximise the revenue from the sale of the by-product mutton, 227.5 units of mutton would need to be disposed of.

The monthly accounts for John Outback would be:

Revenue	Quantity	x	Price	=	Total £s
Wool	46.25	x	36.875	=	1705.4687
Mutton	50	x	10	=	<u>500.00</u>
			Total revenue	=	2205.4687 per month
less Total variable costs equal 92.5 inputs of F x 13.75				=	<u>1271.875</u> per month
			equals Total contribution per month	=	933.5937

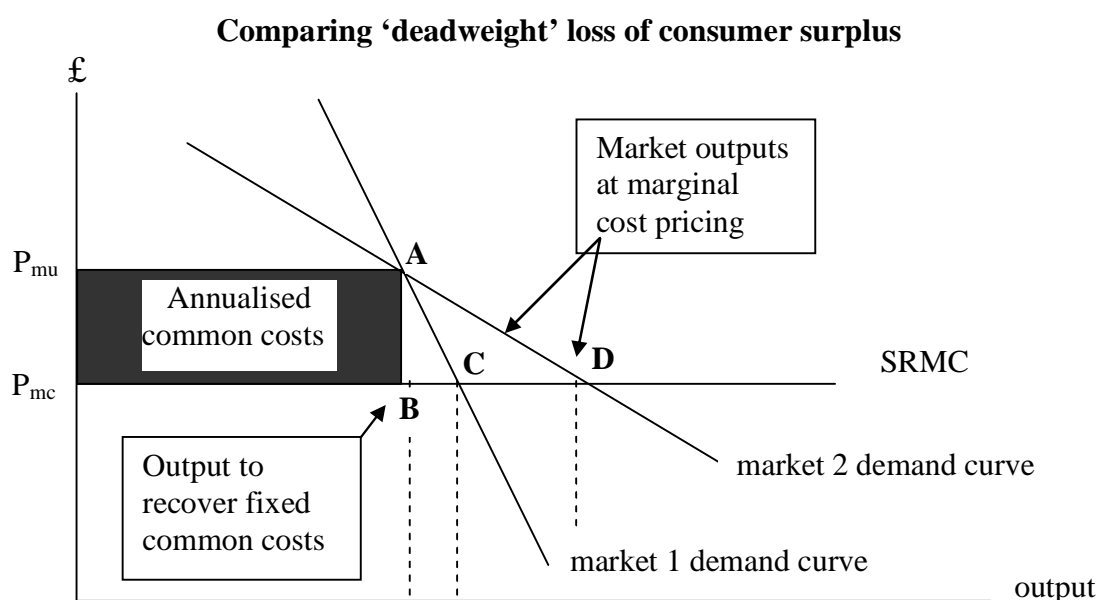
and this amount each month would be the contribution to covering the fixed common costs for the year. As long as these fixed costs were expected to come to less than £11203.124 for the year, John Outback would make an economic profit.

5. Economic allocation of 'common' fixed costs

Economic allocation for joint costs is based on setting the combined marginal revenues equal to the marginal cost of the joint process, but the recovery of 'common' fixed costs is rather different. Common fixed costs do not require products produced to be in fixed proportions, but simply represent those costs which are not variable with the quantity produced, except where the production exceeds the capacity of the asset or activity which is fixed. Fixed costs are important because they often reflect the other two costs of business which have to be added to variable operating costs, being the costs of capital consumption (depreciation) and the cost of capital finance. Setting prices (that is average revenue) equal to marginal costs (ie, those costs

which vary with outputs) may be optimal in the short run, but where there are also fixed costs to be recovered, the revenue earned would be insufficient to ‘break-even’.

One way to recover the fixed common costs across a number of products, where all revenue is gained from selling prices per unit of output, would be to have a common percentage mark-up on marginal cost prices. The problem with this, however, is that each product’s consequent reduction in demand (due to the mark-up causing higher prices) might differ, depending on the demand relationship for each product. This means that the outcome may result in a reduction in ‘consumer surplus’ which is more than it needs to be. The problem can be usefully illustrated from a diagram showing two demand curves for markets which utilise a ‘common cost’. The annualised fixed common costs (area $P_{mu}ABP_{mc}$) could be recovered from either market 1 or market 2 (because their demand curves intersect at point A), but the effect on ‘deadweight’ loss of consumer surplus is clearly different in each market (being either area ACD or area ADB), the two ‘triangles’ being related to the outputs in each market being based on marginal cost pricing alone.



If market 2 was used to recover the common fixed costs the deadweight loss of consumer surplus, area ADB, would clearly be much larger than the deadweight loss if market 1 was used, area ACB. The reason for this is that the demand curve for market 2 at point A is more ‘elastic’ than the demand for market 1 at point A. For the same mark up in price in both markets to recover the fixed costs, it is better to use market 1 because its demand is more inelastic and therefore has a smaller impact on reduced output. Given this, the question arises whether it would be better to recover the common fixed costs from a combination of price increases in both markets because the sum of the deadweight loss of consumer surplus in both markets is less than the loss from using either market alone. A relationship to determine the best combination of price mark ups and quantity reductions in each market to minimise the deadweight loss of consumer surplus was developed by the economist Ramsey and is known as the ‘inverse elasticity rule’.

Inverse elasticity rule

This is normally expressed as an ‘inverse elasticity rule’ for setting optimal mark-ups over marginal cost in each market (ie, set the marked-up prices such that inverse relationship holds). It should be noted, however, that the formula below is expressed in terms of the accountant’s traditional definition of the ‘margin’ (ie, the mark-up as a proportion of the total price). Given

subscript i refers to each particular market, and \overline{P} is the final, marked-up price, and mc is the marginal cost, the margin is defined as $(P_i - mc_i)/P_i$. The inverse elasticity rule, comparing any two markets (1 and 2) is:

$$\frac{\text{margin in market 1}}{\text{margin in market 2}} = \frac{\text{elasticity of demand - market 2}}{\text{elasticity of demand - market 1}}$$

where the margin and the elasticities of demand are based on the outcome, marked-up price in each market. The elasticity of demand is the percentage change in quantity divided by the percentage change in price for a small change in price (which is the 'point' elasticity when Δ is infinitesimally small), that is:

$$\frac{\Delta Q/Q}{\Delta P/P} \text{ or } \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

The inverse elasticity rule achieves the following. If you were to examine the range of price and quantity combinations in the above diagram from all market 1 to all market 2, then as you substituted market 2 for market 1 in small increments, the ratio of margin in market 1 to market 2 would fall, but the inverse ratio of elasticity of demand of market 2 to market 1 would rise. At the point where the curves of the two ratios cross they are equal, and this is the point where the reducing marginal cost of the deadweight loss of consumer surplus in market 1 equals the rising loss of consumer surplus in market 2. This is the point where the sum of the deadweight loss of consumer surplus in both markets is at a minimum, and this is considered a socially optimal way in which to recover the fixed common costs.

The Ramsey pricing rule leads (at least for straight line demand curves) to an equal percentage reduction in quantity demanded in each market, compared with the demand which would otherwise occur with marginal cost pricing. Knowing this, and if we derive the marked-up prices from straight line demand curves, then the Ramsey price in each market will be

$$P_i = a_i - b_i \cdot Q_{mci} \cdot R_q$$

where a_i and b_i are the coefficients of the straight line demand curves for each market, Q_{mci} is the quantity demanded with marginal cost pricing in each market and R_q is the equal proportionate reduction required in each market.

Equal proportionate reductions in quantity supplied

Where there are different proportionate reductions in output (given straight line demand curves), from the level that would apply with marginal cost pricing, then the 'deadweight' loss of consumer surplus from the reduction in output is not minimised. Allowing common cost recovery from equal percentage increases in prices is therefore not socially optimal. Equal proportionate reductions in output minimises the deadweight loss of consumer surplus, but will result in different percentage price mark-ups. The formula for determining the proportionate reduction in quantity supplied in order to recover the common fixed costs is derived from a quadratic equation.

Given the marginal cost per unit is c , the overall percentage reduction $(1-R_q)$ can be calculated from solving a quadratic equation and choosing the appropriate solution from the two which result. This direct solution for Ramsey prices is possible because, given the common fixed costs to be recovered, the only unknown is the overall proportionate reduction R_q . The quadratic can be found from the fact that we must set the sum of total revenues in the markets, less the total variable costs, equal to the common costs to be recovered, that is, $TR^{M1} - TVC^{M1} + TR^{M2} - TVC^{M2} = F$. Given the Ramsey price to be found in each market shown above, then the formula would be for two markets with linear demand curves:

$$(a_1 - b_1 \cdot Q_1 \cdot R_q)(Q_1 \cdot R_q) - Q_1 \cdot R_q \cdot c + (a_2 - b_2 \cdot Q_2 \cdot R_q)(Q_2 \cdot R_q) - Q_2 \cdot R_q \cdot c - F = 0$$

which simplifies to:

$$a_1 \cdot Q_1 \cdot R_q - b_1 \cdot Q_1^2 \cdot R_q^2 - Q_1 \cdot R_q \cdot c + a_2 \cdot Q_2 \cdot R_q - b_2 \cdot Q_2^2 \cdot R_q^2 - Q_2 \cdot R_q \cdot c - F = 0$$

and finally simplifies to the quadratic:

$$- \{b_1.Q_1^2 + b_2.Q_2^2\}R_q^2 + \{(a_1 - c)Q_1 + (a_2 - c)Q_2\}R_q - \{F\} = 0$$

which is solved by the normal quadratic formula for the three bracketed coefficients as shown { }, with the solution taken which is the lowest proportionate reduction R_q from marginal cost based quantities Q_1 and Q_2 . The equal proportionate reduction in quantity in each market, taken as a reasonable approximation in practice from an assumption of straight line demand curves, means that the formal Ramsey pricing rule is much simplified, both for the purposes of understanding and for practical application (eg, it is easy to simulate successive equal proportionate reductions in quantity in each market, with the consequent effects on prices and revenues, and to see when common costs are recovered).

Ramsey prices: a worked example

Fixed common costs of 50 have to be recovered from two markets, which have the following demand curves;

$$\begin{array}{ll} \text{market 1} & : P_1 = 20 - 1Q_1 \\ \text{market 2} & : P_2 = 15 - 0.5Q_2 \end{array}$$

The outputs in each market based on marginal cost pricing, where the marginal cost is 5 per unit of output, are:

$$Q_{mc1} = 15$$

$$Q_{mc2} = 20$$

From the quadratic expression to solve for the required equal proportionate reductions (R_q) in output in each market from the marginal cost pricing outputs above, the coefficients would be:

$$\text{'a'} = -[1(15^2) + 0.5(20^2)] = -425$$

$$\text{'b'} = (20 - 5)(15) + (15 - 5)(20) = 425$$

$$\text{'c'} = \text{common costs to be recovered} = -50$$

These coefficients can then solve for R_q from the standard quadratic formula, choosing the R_q of the two possible solutions which gives the lowest percentage reduction in output:

$$R_q = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}$$

$$\text{ie, } R_q = \frac{-425 \pm 309.23292}{-850}$$

Results

either $R_q = .8638034$, a 13.62% reduction in output from marginal cost levels

or $R_q = .1361965$, an 86.38% reduction in output

we therefore choose $R_q = .8638034$

outputs (Q): market 1 : $R_q Q_{mc1} = .8638034(15) = 12.957051 = Q_1$

market 2 : $R_q Q_{mc2} = .8638034(20) = 17.276068 = Q_2$

prices (P): $P_1 = 20 - 12.957051 = 7.042949$, a mark up of 2.042949

$P_2 = 15 - 0.5(17.276068) = 6.361966$, a mark up of 1.361966

Check to show recovery and the inverse elasticity rule apply

To check that the mark ups recovers the common costs of 50 the following calculations apply:

Recovery = $P_1^m.Q_1 \therefore 12.957051 \times 2.042949 = 26.470594$

$P_2^m.Q_2 \therefore 17.276068 \times 1.361966 = 23.529417$

50.00001 as required for fixed costs.

To check the inverse elasticity rule applies then the following calculations apply:

First calculate the margin in each market from $\frac{P_i - mc_i}{P_i}$

$$\text{Margin}_1 = \frac{2.042949}{7.042949} = .2900701$$

$$\text{Margin}_2 = \frac{1.361966}{6.361966} = .2140794$$

Second, calculate the 'point' elasticity in each market from $\frac{dQ}{dP} \cdot \frac{P}{Q}$

$$\text{Elasticity market}_1 = 1 \cdot \frac{7.042949}{12.957051} = 0.5435611$$

$$\text{Elasticity market}_2 = 2 \cdot \frac{6.361966}{17.276068} = 0.73665062$$

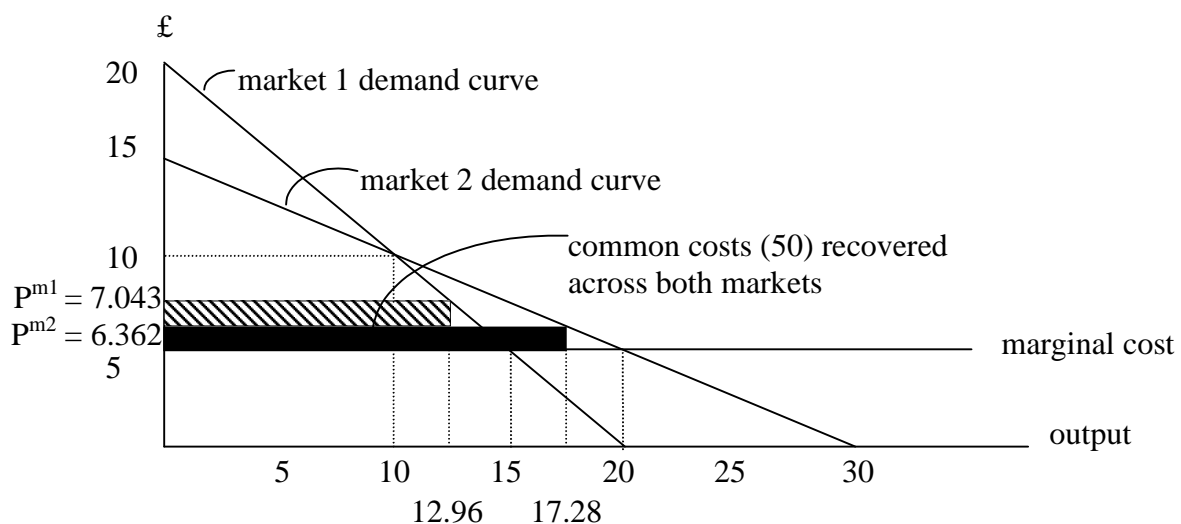
Finally, calculate to show the Ramsey inverse elasticity rule holds for the equal proportionate reductions in output

$$\text{Ratio of margins is: } \frac{.2900701}{.2140794} = 1.354965 = \mathbf{1.355}$$

$$\text{Inverse ratio of elasticities is: } \frac{0.73665062}{0.5435611} = 1.3552305 = \mathbf{1.355}$$

Both ratios equal 1.355 therefore the relationship holds. The outcome of the above is shown diagrammatically below, where it can be seen that the most inelastic market has the highest mark-up.

The 'Ramsey' price outcome



with the Ramsey prices rising from the marginal cost of 5 to 7.043 and 6.362 in order to recover the fixed common costs of 50 and with the equal proportionate reduction in quantity for each market of 13.6% from 15 to 12.96 and from 20 to 17.28.

Comparing Ramsey pricing with other alternatives

The lost consumer surplus from the increase in prices and reduction in quantity supplied in order to recover the fixed common costs using Ramsey pricing can be compared with alternatives which do not have an equal proportionate reduction in output. The following examples include equal percentage mark-ups in price over marginal cost, and the two options all market 1 or market 2. The effects on deadweight loss of consumer surplus can then usefully be shown in a table. To calculate the effect of an equal percentage mark up in price to recover the fixed common costs we need to use a quadratic in a similar way to the calculation for an equal proportionate reduction in quantity.

* Deadweight loss with constant mark-up in price

In this case, instead of solving a quadratic to find the constant reduction in quantity, we need one to solve for the equal increases in prices. Given the initial prices in both markets would be based on the marginal cost (c), then if (m) is to be the constant mark-up multiplier, then P_1 and

P_2 have to be equal to $(c.m)$. As before, we need total revenue minus total variable costs minus total fixed common costs to be recovered to equal zero. Given the two demand curves, Q_1 and Q_2 can be expressed in terms of P_1 and P_2 (for example, if $P = a - bQ$, then $Q = (P - a)/-b$. As before the quadratic can be found from simplifying and rearranging, and the quadratic for the two markets would be:

$$\{(1/-b_1).c^2 + (1/-b_2).c^2\}m^2 - \{(1/-b_1).c(a_1+c) + (1/-b_2).c(a_2+c)\}m + c\{(1/-b_1).a_1 + (1/-b_2).a_2\} - \{F\} = 0$$

The result of solving this quadratic would be that m could equal either 1.33333 or 3. A mark-up of 1.33333 times the marginal cost price would clearly be the one to choose, and the result would therefore be a price in both markets 1 and 2 of $5 \times 1.33333 = 6.66665$. The quantity in each market would be $Q_1 = 13.333334$ (an 11.111% reduction) and $Q_2 = 16.666665$ (a 16.666% reduction).

The table is in three parts for the four options to be compared. First, quantity comparisons, then price comparisons, and finally the net revenue (TR – TVC) and deadweight loss of consumer surplus comparisons:

Quantity	Qmc	mkt1 only	mkt2 only	constant mark up	Ramsey pricing
Market 1	15	10 (-66.6%)	15 (0%)	13.33 (-11.13%)	12.96 (-13.6%)
Market 2	20	20 (0%)	10 (-50%)	16.66 (-16.7%)	17.28 (-13.6%)
Price	Pmc				
Market 1	5	10 (+100%)	5 (0%)	6.667 (+33.34%)	7.043 (+40.86%)
Market 2	5	5 (0%)	10 (+100%)	6.667 (+33.34%)	6.362 (+27.24%)

Revenues	Total revenue	Total variable cost	Net revenue	Deadweight loss of CS
Mkt1 only	1. 10x10=100 2. 20x5 = <u>100</u> 200	10x5=50 20x5= <u>100</u> 150	50 <u>0</u> 50	(15-10)(10-5)÷2 = 12.5
Mkt2 only	1. 15x5 =75 2. 10x10= <u>100</u> 175	15x5=75 10x5= <u>50</u> 125	0 <u>50</u> 50	(20-10)(10-5)÷2 = 25
MU%	1. 13.33x6.667=88.871 2. 16.66x6.667= <u>111.072</u> 199.943	13.33x5=66.65 16.667x5= <u>83.335</u> 149.985	22.221 <u>27.737</u> 49.958	(15-13.33)(6.667-5)÷2=1.392 (20-16.66)(6.667-5)÷2= <u>2.784</u> 4.176
Ramsey	1. 12.96x7.043=91.278 2. 17.28x6.362= <u>109.935</u> 201.213	12.96x5=64.8 17.28x5= <u>86.4</u> 151.2	26.478 <u>23.535</u> 50.013	(15-12.96)(7.043-5)÷2=2.084 (20-17.28)(6.362-5)÷2= <u>1.852</u> 3.936

This table shows that of the four options Ramsey pricing has the lowest deadweight loss of consumer surplus at 3.936, and although a constant mark up in price in each market is better at 4.176 than recovering the common fixed costs from one market alone (25 or 12.5), it is still better to recover them from equal reductions in quantity. The table also shows that each of the four price/quantity options is sufficient to recover the common fixed costs of 50. Having looked at the four outcomes, the following table shows the deadweight loss in consumer surplus for each option and the percentage points deviations from Ramsey pricing quantities for the other options.

	Mkt1 % Reduction	Mkt2 % Reduction	Deadweight loss of consumer surplus	% points deviation from Ramsey pricing
Ramsey pricing	13.6%	13.6%	3.936	0
Mark-up constant	11.13%	16.7%	4.176	+2.3/-3.1
Option 1	66.6%	0%	12.5	-53/+13.6
Option 2	0%	50%	25	+13.6/-50

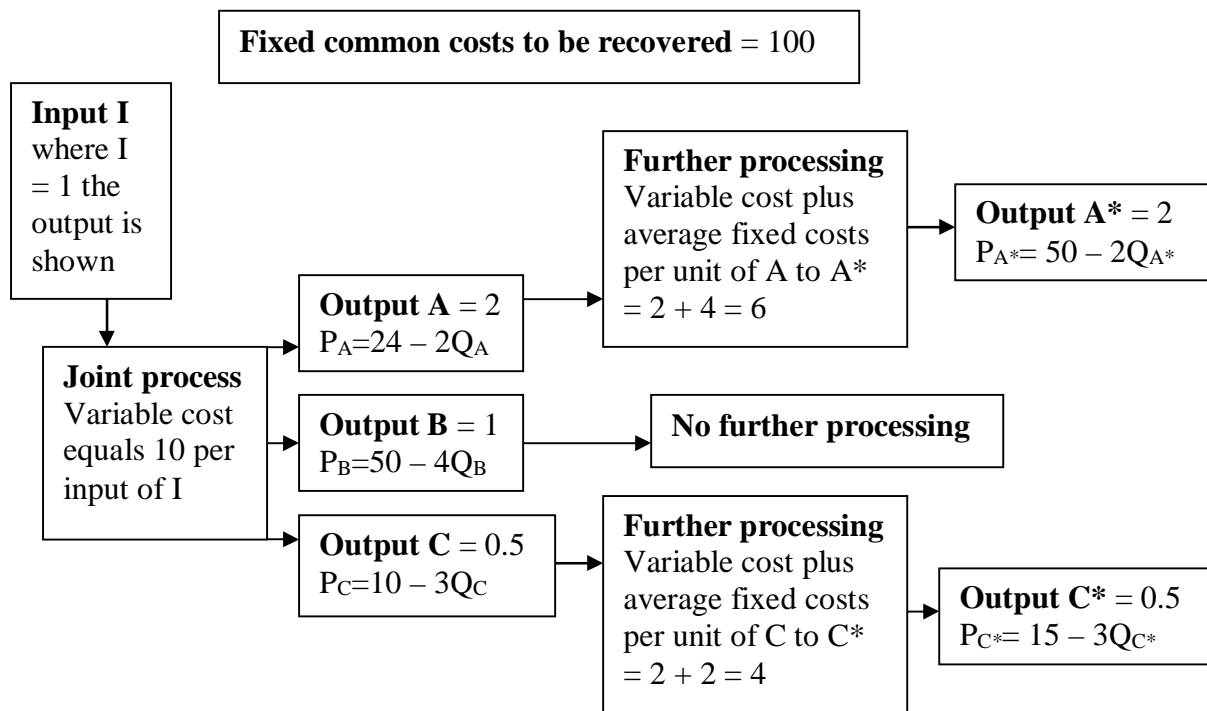
The use of Ramsey pricing has also been an important part of the debate on tax policy, given that if a government wants to raise a fixed amount of tax from a levy on goods and services, rather than having a constant mark-up on each sale price, the cost of lost consumer surplus would be minimized by having a tax system which resulted in equal proportionate reductions in the quantities demanded.

Combining joint processes with further processing and common fixed costs

The previous two examples are interesting because joint process optimisation was based on setting marginal revenue equal to marginal cost ($MR=MC$) in order to maximise contribution, and that contribution may be sufficient to recover any fixed costs. Ramsey pricing however is based on recovering fixed costs given marginal cost pricing, that is setting average revenue equal to marginal cost ($AR=MC$). Marginal cost pricing based on short run variable costs is therefore unable to recover the other two costs of business involved in fixed costs (depreciation and the cost of capital finance), and prices therefore need to rise above marginal cost in order to recover them. The relevance of marginal cost pricing was seen to be very important for government policy on nationalised industries, particularly the utilities and network industries. Given sufficient capacity, it was argued that welfare would be increased by pricing on variable costs, given there is no cost to using existing capacity. However, assets need to be replaced, and the cost would then have to be financed. This could be directly from taxation, but for the most part the government lends the money to the utility. Given this, if prices were based on variable costs then the debt to government would build up and eventually have to be written off. It is difficult to maintain a system which is not based on prices which cover at least the three costs of business, given concerns over equity between consumers and taxpayers, and intergenerational equity, and the debate changed anyway when utilities and network industries were privatised. Another important issue that arises is when capacity is reached; should prices rise to limit demand? Congestion charging on roads is a good example.

Given competition should drive down prices towards the point where total revenue equals total cost, then average revenue and average costs are an important part of managerial decisions. The following example examines the situation where a decision has to be made on the amount of a joint process to be run in a period when there are different periodic demand curves which can arise from further processing, and fixed common costs have to be recovered. The following diagram illustrates the production process. The diagram shows the output/input relationships from the joint process, being 2 for A, 1 for B and 0.5 for C for each unit of I, and that with further processing A can be converted into A* with a different demand curve, and similarly for output C. The first step is to ask whether A* is to be preferred to A, and whether C* is to be preferred to C? The difference between the two demand curves is the incremental gain in average revenue from further processing, and so these can be compared with the costs of the further processing. For A, the average revenue increase is 26 (ie, 50 – 24) and this is greater than the further processing costs of 6, so A* is preferred to A. Similarly for B, the average revenue increase is 5 (ie, 15 – 10) and this is greater than the further processing costs of 4, so C* is preferred to C. The average fixed costs of further processing are included at this stage because the costs relate specifically to a particular output, unlike the common fixed costs.

Combining Fixed common costs, joint processes and further processing



Having chosen A*, B and C* as the products to be achieved from the production process, the next step is to decide the input of I necessary to maximise production before recovery of fixed common costs. The first stage is to see whether any of the outputs A, B and C are 'by-products'. This requires us to set the sum of the average revenues of A*, B and C equal to variable cost of the joint process, taking account of the output/input relationships. As A equals 2I, then $50 - 2Q_{A*}$ becomes $50 - 4I$; as B equals 1I, then $50 - 4Q_B$ becomes $50 - 4I$; and as C = 0.5I, then $15 - 3Q_{C*}$ becomes $15 - 1.5I$. This enables us to solve for I, given the variable cost of the joint process equals 10. Setting $50 - 4I + 50 - 4I + 15 - 1.5I = 10$ we get $I = 11.052631$. For that input of I the average revenue of A* at that point is 5.79, as it is for B. However, for C* at that point it is -1.579, so C* is a by-product. Eliminating C* from the equation, the optimal input of I increases and $I = 11.25$. At that point the average revenue of A* is 5, as it is for B. The sum of the average revenues equal 10, so in effect the variable costs of the joint process have been allocated to A* and B proportional to their average revenue at that point, and no joint process costs are allocated to C*. At an input of 11.25, the output of C = 5.625 but the question then is how much of this is to be further processed into C*. Given no joint process costs have been allocated to C then the answer is to set $15 - 3Q_{C*} = 4$, the further processing costs, and the answer is $Q_{C*} = 3.6667$, meaning 1.9583 units of C remain unused.

The further processing costs of A* also have to be recovered, and as the price of A* rises to recover it, then the amount of A being used will fall, and this means the allocated joint process costs to A* will not be fully recovered. The amount of joint process costs allocated to A* is $11.25 \times 10 \times 0.5 = 56.25$, and these need to be recovered along with the further processing costs. The equation necessary is to set the total revenue of A* - joint process costs - further processing costs = 0. This would be $(50 - 2Q_{A*})Q_{A*} - 56.25 - 6Q_{A*} = 0$. Multiplied out and rearranged this gives a quadratic of $-2Q_{A*}^2 + (50 - 6)Q_{A*} - 56.25 = 0$. Solving this by the normal quadratic formula, the answer is that Q_{A*} can be either 20.637167 or 1.3628325. Clearly we would choose 20.637167, a reduction of 1.862833 from the production of A from the joint

process of 22.5 (ie, 11.25 x 2). At this stage A* has in effect become a form of by-product because it is not using all of A from the joint process. Only B is using its full 11.25.

At this stage the results are that:

$$\begin{array}{llll} I = 11.25 & Q_{A^*} = 20.6372 & Q_B = 11.25 & Q_{C^*} = 3.6667 \\ & P_{A^*} = 8.7257 & P_B = 5 & P_{C^*} = 4 \\ & \text{Unused A} = 1.8628 & & \text{Unused C} = 1.958 \end{array}$$

Given this information, the last stage is to recover the common fixed costs by Ramsey pricing, a common percentage reduction R in quantity from the current quantities for A*, B and C* above to raise the prices to the necessary level. The formulae for each of A*, B and C* will incorporate their various costs and revenues, and the cost of the joint process can fall to B because only it is using its full output from the joint process. The result would be:

$$\begin{array}{lll} \text{For A*} & 50RQ_{A^*} - 2(RQ_{A^*})^2 - 6RQ_{A^*} & = \text{TR} - \text{FP of A to A*} \\ \text{For B} & 50RQ_B - 4(RQ_B)^2 - 10RQ_B & = \text{TR} - \text{JP costs} \\ \text{For C*} & 15RQ_{C^*} - 3(RQ_{C^*})^2 - 4RQ_{C^*} & = \text{TR} - \text{FP of C to C*} \end{array}$$

the sum of which must equal the fixed common costs of 100.

After substituting the values for Q_{A^*} , Q_B and Q_{C^*} in the above formulae, rearranging and simplifying, the quadratic is $-1398.3271R^2 + 1398.3705R - 100 = 0$. Solving this would give a value of R which is either 0.9224774 or 0.07752. Choosing the lowest percentage reduction would yield a figure of $1 - 0.9224774$, a 7.75226% reduction in quantities. The result is that:

For A* the quantity becomes $20.637167 \times 0.9224774 = \mathbf{19.03732}$ with $P_{A^*} = \mathbf{11.92536}$
 For B the quantity becomes $11.25 \times 0.9224774 = \mathbf{10.37787}$ with $P_B = \mathbf{8.48852}$
 For C* the quantity becomes $3.38236 \times 0.9224774 = \mathbf{3.38236}$ with $P_{C^*} = \mathbf{4.85292}$

The 'final' accounts would be:

	A*	B	C*	Σ
Total Revenue	227.02689	88.092757	16.414322	331.53396
less further processing	114.22392	0	13.52944	- 127.75336
less joint processing		103.7787		- 103.7787
less joint common costs				- <u>100.0</u>
equals 'Break even' (ignoring the small rounding error)				0.0019

The allocation of the costs above in terms of prices is as follows:

	P_{A^*}	P_B	P_{C^*}
Joint process costs	2.726	5	0
Further processing costs	6.0	0	4
Joint common costs	<u>3.199</u>	<u>3.489</u>	<u>0.853</u>
equals final price	11.925	8.489	4.853

The end result is to produce enough in the period which can absorb the costs and have an overall outcome where total revenue equals total cost, and the deadweight loss of consumer surplus to recover fixed costs is kept to a minimum. Some of the output of the joint process is unused in further processing for A* and C*, but if there was a disposal cost of 0.6 per unit of A remaining and 0.3 per unit of C remaining to be recovered because of this, then it would have to be added into the Ramsey pricing formulae, perhaps by defining common costs to include the unused amounts of A and C which would be $100 + (RQ_B - RQ_{A^*}) \times 0.6 + (RQ_{C^*} - RQ_{C^*}) \times 0.3$

6. Economic depreciation and expected technical change

Depreciation is the allocation of the 'joint cost' (with respect to time) of a capital asset over the life of the asset, and ensures that the reducing value of the asset is charged as an expense to the profit and loss account over the life of the asset. The profile of the depreciation charge for each year over the life of the asset can vary between an equal amount each year (straight line depreciation) to those which either charge more depreciation expense in early years (accelerated or 'front-loaded') or less depreciation expense in early years (deferred or 'back-loaded'). Whichever is chosen should, however, relate to the economic value of the asset, since it is not good accounting practice to have a net book value for an asset in the balance sheet which is higher than its realisable value for that year.

If the asset is retained in the business, in one sense the allocation is arbitrary, given it is a joint cost over time. However, capex is not a sunk cost, given the potential of its realisable value. The realisable value is the 'opportunity cost' of holding the asset in the business and, therefore, depreciation can be linked to 'economic reality'. The economic measure of depreciation in the year would therefore be the 'using-up' of avoidable cost measured by realisable value at each year end. The economic depreciation schedule (EDS) would therefore be:

$$\text{year 1} = \text{NRV}_{10} - \text{NRV}_{1C} = \text{EDS charge (year 1)}$$

$$\text{year 2} = \text{NRV}_{20} - \text{NRV}_{2C} = \text{EDS charge (year 2)}$$

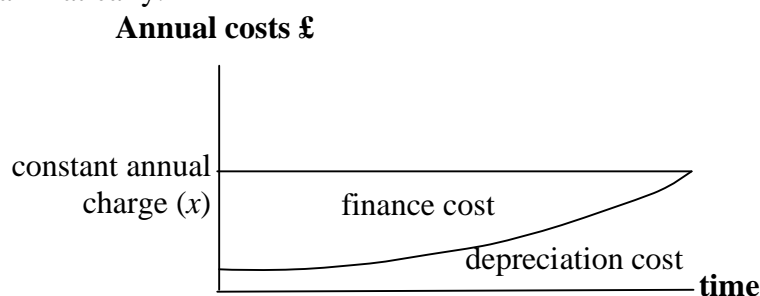
: and so on :

This type of schedule, other things being equal, fits well with the underlying depreciation schedule of the constant annuitised charge for a capital asset. The underlying depreciation schedule is illustrated as follows. Given the investment $I = 100$, it has a two year life, and a cost of capital = 10%, then set $100 = x(.909) + x(.826)$

and therefore the constant annual charge $x = 57.6\text{pa}$. The split of depreciation and cost of finance, given the constant annuitised charge, would be, given return on investment until the return of investment by way of depreciation:

	NBV ₀	Finance charge	Depreciation	NBV _c
Year 1	100	(100 @ 10%)=10	(57.6 - 10) = 47.6	52.4
Year 2	52.4	(52.4 @ 10%)=5.24	(57.6 - 5.24)=52.36	0.04 (rounding error)
		15.24	100	

The rising depreciation charge and falling finance charge each year can be shown diagrammatically:



Given the annuity formula: $I = x(df_1) + x(df_2) \dots x(df_n)$ then the terms on the right hand side can be seen as an allocation of the cost of the investment over the life of the asset (ie, a depreciation schedule). Other things being equal, what would be the realisable value each year for a three year asset, given the revenue which could be earned from the asset was x pa? Given years 2 and 3 are

looked at from the point of view of the potential purchaser being at the start of years 2 and 3 respectively, it would be:

$$NRV_{10} = x(df_1) + x(df_2) + x(df_3)$$

$$NRV_{20} = x(df_1) + x(df_2)$$

$$NRV_{30} = x(df_1)$$

Clearly, the depreciation charge schedule based on differences in NRV_0 for each year would be:

$$\text{Year 1 (ie, } NRV_1 - NRV_2) = x(df_3)$$

$$\text{Year 2 (ie, } NRV_2 - NRV_3) = x(df_2)$$

$$\text{Year 3 (ie, } NRV_3 - NRV_4) = x(df_1)$$

which is simply the annuity allocation reversed. The annuity depreciation schedule therefore links in with the 'economic' depreciation concept (given the assumptions made). Note the relationship between each year's depreciation charge, given r is the firm's cost of capital:

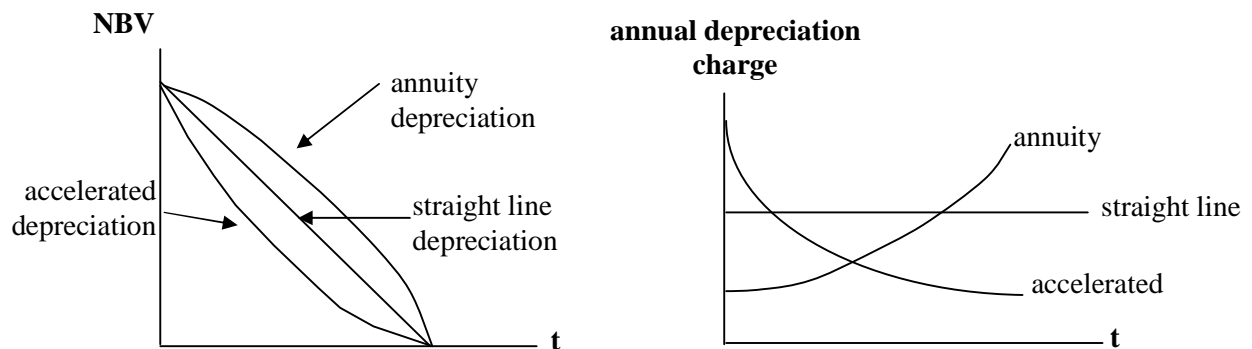
$$x(df_2) = x(df_3)(1 + r)$$

$$x(df_1) = x(df_3)(1 + r)^2$$

so that the annuity depreciation schedule can be seen as a 'compound' depreciation schedule based on a first year depreciation charge of $x(df_n)$, and where $x(df_n) = x - Ir$, as used in our worked example schedule. To show this, using the figures in the example:

$$47.6(1.1) = 52.36.$$

The relationship between the three basic types of depreciation schedule for individual assets can be illustrated as follows:



From the point of view of financial control, the most important consideration is to forecast a depreciation schedule which reasonably matches the decline in realisable value that actually occurs, otherwise the organisation's auditors will require large 'impairment' write-offs at some time in the future because your assets cannot realise their current depreciated net book value. This is particularly a problem with 'high tech' industries, where technological advances from competitors (and hence lower cost-reflective prices) may mean that your past investments become sunk costs which cannot be recovered. Hence prudent managers front-load depreciation schedules where there is exposure to technological risk of new processes and products in order to recover past investments while they can, balanced by the need to set prices competitively in current markets.

Managing depreciation schedules with technical change

The profile of 'economic' depreciation charges is determined by the decline in the realisable value of an asset over time. In a certain, steady state world (ie, without technical change or other efficiency improvements), the compound annuity depreciation schedule would reflect the decline in the realisable value. This might seem surprising, given the emphasis in accounting and business practice on straight line depreciation (or more rapid depreciation schedules). In one respect, this can be explained by the real world being uncertain (and hence it is prudent to recover your investment while you can) but it can also be observed that straight line depreciation

typically derives from the apportionment of only one element of the cost of capital assets, that is, the recovery of the acquisition cost. There is, however, the other cost, the cost of capital finance. An asset has output serviceability over the whole of its life and, other things being equal, it is the equal apportionment of the cost of both elements of the cost of capital assets which one would expect to drive the realisable value of an asset over its life. This is what the annuitised constant annual charge achieves.

Figure 1: traditional depreciation

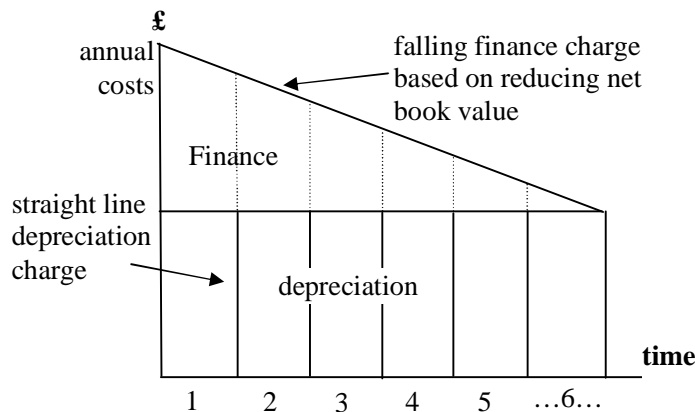
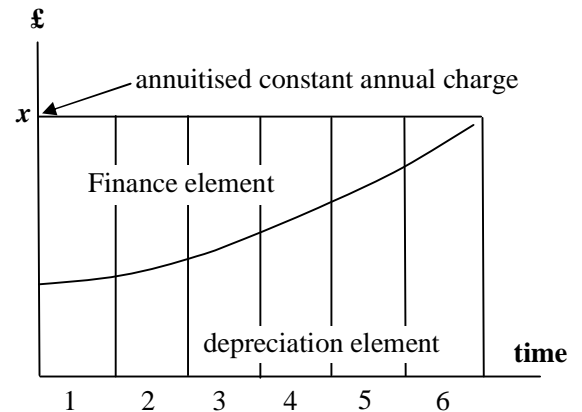


Figure 2: compound annuity depreciation



Diagrammatically (using a six year life asset as an example), this is saying that **Figure 2** has ‘economic’ foundations related to expected realisable value, compared to **Figure 1**.

Prices and technical change

However, if technical and other efficiency improvements reduce the acquisition cost of new capital assets over time, then competitive market prices for existing asset owners will be determined by the lower costs facing potential ‘new entrants’. The realisable value of existing capital assets will fall to reflect that, and financial controllers therefore have to anticipate this in their management of the depreciation profile. If prices are going to be forced lower in the future, then prices are going to have to be higher now (ie, before the new technology comes in) in order to recover the costs of the investment. In practice, this means ‘tilting’ the depreciation towards the present, in line with the anticipated profile of technical improvement, and hence falling real prices. This can be illustrated as follows:

Assume technical change occurs at a price reduction rate (θ) each year (eg, 5% improvement would give a θ factor of .95). This would mean the acquisition cost of an investment (I) now would be $I.\theta$ next year and $I.\theta^2$ the following year and so on. Equivalently, each year, looking forward from that point, the constant annuitised charge (x) for a new asset would fall by θ , θ^2 ,... and so on. Note that competitive prices are ‘cost-reflective’, so that charge x is interpreted as the price which needs to be set in the market to recover the cost of the capital investment (capital consumption and capital finance). In this case therefore x can be replaced by P .

To take account of anticipated technical change, a financial controller would manage the depreciation schedule, and hence the cost-reflective pricing, by setting out the following equation (using the example of a three year life asset) in exactly the same way as the constant annuitised charge was calculated:

$$\text{Set acquisition cost } I = P_1.df_1 + P_2.df_2 + P_3.df_3$$

$$\text{where } P_2 = P_1.\theta$$

$$P_3 = P_1.\theta^2$$

This means that the base reference price \bar{P}_1 can be solved for, as P_2 and P_3 are a function of P_1 , given the rate of technological change θ . Given:

$$I = \bar{P}_1(df_1) + \bar{P}_1.\theta(df_2) + \bar{P}_1.\theta^2(df_3)$$

then
$$\bar{P}_1 = \frac{I}{df_1 + \theta(df_2) + \theta^2(df_3)}$$

The consequential depreciation schedule

The terms of the above expression represent the present values of the recoverable amounts of the asset in each year, given the competitive price profile anticipated due to technical change. The resultant depreciation charge each year can be found from the expected decline in the realisable value of the asset (RVI). This can be shown as:

acquisition cost I = $RVI(3) = \bar{P}_1.df_1 + P_2.df_2 + P_3.df_3$

: for the asset with three years life left

next year $RVI(2) = P_2.df_1 + P_3.df_2$

: the realisable value with two years left

following year $RVI(1) = P_3.df_1$

: the realisable value with one year left

The ‘economic’ depreciation profile is given by the decline in the RVI, therefore the differences between them will be the annual depreciation charges:

Year 1 depreciation = $(\bar{P}_1 - P_2)df_1 + (P_2 - P_3)df_2 + P_3.df_3$

Year 2 depreciation = $(P_2 - P_3)df_1 + P_3.df_2$

Year 3 depreciation = $P_3.df_1$

This can be illustrated as shown in **Figure 3** (the figures shown are drawn from the following worked example), noting in particular that the last term in each year’s depreciation formula shown above is the compound annuity depreciation schedule (as found with the constant annuitised annual charges). The other terms reflect the write-offs of acquisition cost related to the ‘impairment’ of the asset’s realisable value due to falling prices over time resulting from technical change. **Note** too, using year 1 depreciation as an example, how this divides into two further distinct components of depreciation. If we divide the third term $P_3.df_3$ into two parts, ie:

$$P_3.df_3 = (P_3 - P_4) df_3 + P_4.df_3$$

then we have the first three terms for year 1 depreciation above as:

$$(\bar{P}_1 - P_2)df_1 + (P_2 - P_3)df_2 + (P_3 - P_4)df_3 \text{ plus } P_4.df_3$$

The first three terms can now be seen as simply the fall in the gross acquisition cost of the asset over the first year because:

gross cost now of I = $\bar{P}_1.df_1 + P_2.df_2 + P_3.df_3$

gross cost next year of I = $P_2.df_1 + P_3.df_2 + P_4.df_3$

The last term ($P_4.df_3$), as with the constant annuity charge, is the first year of a compound annuity depreciation schedule, but for an asset taking into account the reduction in asset prices due to technical change. In short, the first year’s depreciation charge is the write-off of all of the reduction in the gross acquisition cost of the asset due to technical change, plus the appropriate annual annuity depreciation charge for the year, but based on the new, lower acquisition cost, given the write-off. This is:

$(I_1 - I_2) + P_4 \cdot df_3$. **Note too**, that if technical change is zero, so that $\theta = 1$, then $P_1 = P_2 = P_3 = x$, the original constant annuity charge, and the first term becomes zero, leaving the standard result for the compound annuity depreciation schedule.

Worked example: A three year asset costing 100; $r^* = 10\%$; $\theta = 5\%$ pa

	1	2	3
discount factors	.9091	.8264	.7513
technology factors	.95	.9025	.8574

The base reference price is:

$$\bar{P}_1 = \frac{100}{.9091 + (.95)(.8264) + (.9025)(.7513)} = \frac{100}{2.3722} = 42.155$$

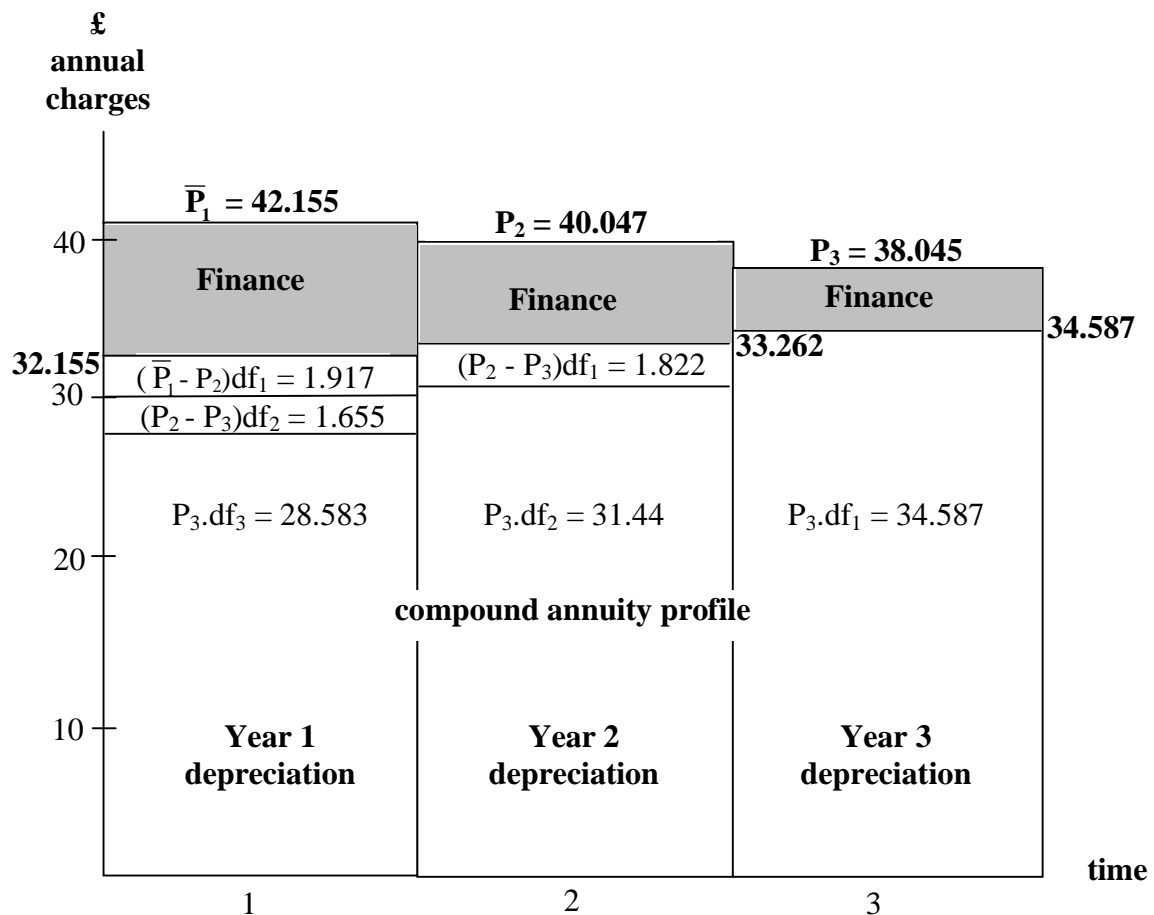
The price profile is therefore:

$$\begin{aligned}\bar{P}_1 &= 42.155 \\ P_2 &= (42.155)(.95) = 40.047 \\ P_3 &= (42.155)(.9025) = 38.045 \\ P_4 &= (42.155)(.8574) = 36.144\end{aligned}$$

The depreciation schedule can be calculated as:

	total annual charge	interest on NBV _o	depreciation	NBV _c
Year 1	42.155	10	32.155	67.845
Year 2	40.047	6.785	33.262	34.583
Year 3	38.045	3.458	34.587	- 0.004
			100.004	rounding error

Figure 3: the 'tilted' depreciation schedule



Note that the depreciation schedule is still ‘backloaded’, ie rising over time, and can be compared with the constant annuitised charge (x), ie, constant annual price with no technical change (ie, $\theta = 1$). This would be:

$$x = 40.21 = \bar{P}_1; P_2; P_3; P_4 \dots \quad \text{Depreciation: year} \quad \begin{array}{l} 1 = 30.21 \\ 2 = 33.23 \\ 3 = 36.55 \end{array}$$

Compare the depreciation charge for year 1 derived from the above schedule (32.155) with the formula depreciation charge derived above:

$$\begin{aligned} & (I_1 - I_2) + P_4(df_3) \\ &= (I - I.\theta) + \bar{P}_1.\theta^3(df_3) \\ &= 100(1 - .95) + (42.155)(.8574)(.7513) \\ &= 5 + 27.155 \\ &= 32.155, \text{ as above} \end{aligned}$$

The comparison shows that compared with the standard annuity depreciation schedule, the combination of annuity depreciation with expectations of impairment of asset value due to technical change makes the annual depreciation charge move to more of a straight line. The last aspect which tends to move it further towards an ‘accelerated’ depreciation profile is the issue of rising maintenance costs over the life of the asset.

Rising maintenance costs, realisable value and the depreciation profile

Suppose maintenance costs rise over the life of an asset, such as a car. This makes it necessary to compare the underlying market price, which is based on the annuitized present value of the maintenance profile over the full asset life with the annuitized value arising from the remaining life for someone buying a ‘second hand’ asset. The following example illustrates this based on a three year asset with a cost of capital of 10%, and with the yearly maintenance costs of:

Year	1	2	3
Maintenance costs	10	15	25

For a new car the present value of the series of maintenance costs would be:

$$10(.909) + 15(.826) + 25(.751) = 40.255$$

and its ‘annuitised’ value for the annual charge to cover it would be 40.255 divided by the sum of the three discount factors, being 2.487. The annual charge would be 16.1826.

What deduction would a buyer expect from the value of the asset if they were buying it one year old? The present value of the remaining maintenance schedule would be:

$$15(.909) + 25(.826) = 34.285$$

and the annuitised annual charge would be $34.285/1.736 = 19.7494$.

Given the recoverable value from market prices for new cars is only 16.1826, then, in effect, the 3.5632 difference is ‘unrecoverable’. The present value for the remaining two years is therefore $3.5632 \times 1.736 = 6.186$, and this would be deducted from the expected realisable value of the asset based on its replacement cost. This deduction is necessary in order to make the purchaser indifferent between buying a new asset and a second hand asset, other things being equal. If the realisable value has fallen then the depreciation charge for the previous year has to rise.

After two years the deduction becomes even higher, given the rising maintenance costs. The present value of maintenance at the start of year three would be $25(.909) = 22.725$, but only $16.1862(.909) = 14.713$ could be recovered from the market. The amount offered for the two year old asset would therefore have to be reduced by $22.725 - 14.713 = 8.012$.

The conclusion has to be that accountants' straight line or accelerated depreciation schedules are likely to be correct from an economics point of view, given they are made up of:

- * Base annuity depreciation
- * Impairment related to falling replacement prices
- * Deductions for rising maintenance costs over the life of the asset

The following table is a summary of the worked examples, showing how as each element is added in the depreciation profile becomes more 'front' loaded:

Component **annuity depreciation** **plus** impairment **plus** maintenance abatement

NBV/RV	Yr 0	100	100	100
	Yr 1	69.79	67.84	61.65 (-6.19)
	Yr 2	36.55	34.59	26.58 (-8.01)
	Yr 3	0	0	0
Depreciation				
charge pa	Yr 1	30.21	32.16	38.35
	Yr 2	33.24	33.25	35.07
	Yr 3	<u>36.55</u>	<u>34.59</u>	<u>26.58</u>
		100.00	100.00	100.00
Depreciation				
Profile		deferred	more	accelerated
			straight line	

IV Independent Economic Regulation - Public Services Privately Provided

The privatisation and regulation of the utility and network industries (water, energy, transport and communications) in the UK between 1980 and 2013 provide a classic example of political economy in practice. Between 1945 and 1979, these industries had been nationalised and public opinion, or at least expectation, had been that, in general, these industries should be the responsibility of the state. The development of 'independent' economic regulation alongside the privatisation of each of the industries was an important fusion of government policy, the politics involved, economic method, accounting practice, and new institutional frameworks. Change occurred as experience built up, and particularly when the government changed at elections, notably 1979, 1997 and 2011.

The purpose of part IV is to provide an overview set over seven sections: 1. Why regulate?, 2. Nationalisation and privatisation, 3. Types of competition; 4. The role of independent regulators, 5. The 'Littlechild' method (generally known as $RPI - x$), 6. The economic history 1984 – 2013, and 7. An overview and bibliography. A bibliography is included for those who wish to look at independent economic regulation as a specialist subject.

1. Why regulate?

At its heart the question is really 'why government?', and the answer is clear, because the citizens of any country need a government to take responsibility for those many things which would not be effective, or efficient, if citizens tried to carry them out for themselves. Nevertheless, when it comes to the way governments approach market failure, there can be many questions as to 'why regulate?' in particular circumstances.

Political economy has been extensively applied in the area of how to provide and control the essential service 'network' industries, covering water, energy, transport and communications. This is because they either have a large degree of monopoly power, which could be abused, or because the infrastructure tends to have a degree of 'natural' monopoly, which means the 'efficient' solution should only have one supplier. The importance of these industries naturally gives them a high political profile, but such a high political profile can of itself cause problems and barriers to achieving sensible outcomes. This is particularly notable when it comes to the question of whether such industries should be publicly or privately owned. Some would argue that such important industries must be publicly owned, and that state employees in these industries will imbibe a 'public service' ethic which means it will always be run in the public interest. Others would argue that the role of government is to ensure effective outcomes which meet their democratically endorsed objectives and purposes. If those outcomes can be better achieved by providers and suppliers from the private sector, given effective regulation, then there seems to be no problem in principle of having 'public services privately provided'.

National defence is clearly a public service, but it is self-evident that many of the people involved in defence services are not necessarily state employees, but may be operating under contract to the state. Military uniforms and hardware, for example, can be supplied by private sector, and those suppliers may also be supplying abroad, exporting into the international market place. It is no different in many other services, whether they are catering or consulting services. The regulation of contracted services has to be a key balance in public services, privately provided, and should not fall foul of the general opinion that regulation is a 'bureaucratic nightmare' and therefore regulation is intrinsically a bad thing. Unnecessary regulation clearly is, necessary regulation clearly is not.

The experiment in political economy represented by the privatisation and regulation of the utility and network industries from 1984 in the UK is an important example of how government can retain its fundamental purpose but carry out the delivery of the required outcomes in a different way. That privatisation of these industries should have been accompanied by regulation was inevitable. For the most part they were state monopolies, and as Adam Smith said in the *Wealth of Nations*: “People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public or in some contrivance to raise prices”.

2. Nationalisation and privatisation

Government regulates industries whether they are publicly or privately owned. The model which was adopted for regulating these ‘infrastructure’ industries on privatisation was designed to meet two key objectives, first, ensuring that ‘all reasonable demands’ for supply would be well met, and second ensuring that sustainable finance and cost-effective pricing were in place. These were to ensure that the consuming public would be supportive of the policy of privatisation while at the same time the investors would have the long term confidence to buy the shares offered for sale. These had to be addressed because while the industries were nationalised government was often faced with a conflict of interest between ownership and regulation, and the regulation of the financial side could be affected by the short run politics which affect debates on public expenditure.

The ‘infrastructure’ industries had to finance their capital expenditure by borrowing from the government, and had to receive political approval for the capital expenditure programmes as part of the public expenditure review process. Capital finance was repaid to the government over the long term in terms of principal repayments and interest. Two problems can evidently arise in such long term industries. Maintenance which should be done can be deferred (even though it cannot be deferred indefinitely), thereby reducing public expenditure in the short run, and if it is politically useful to the government to keep customer prices low in these industries, the government can lend money to them to finance any deficit arising from prices being held lower than cost. In the longer run it was found that such debts might have to be written off by the government. The problems were evident at the time of privatisation, because investors buying the industries found that they might have to factor in a large surplus of deferred maintenance, and customer prices which were below the level of cost for a ‘going concern’ in the long run. The result was that the share price for selling the nationalised industries was much lower than many expected.

The solution was to create new regulatory bodies at the time of each privatisation, thereby making the regulation separate from the politicians and the owners of the shares. The new ‘independent’ economic regulators were ‘Non-ministerial public bodies’ (NMPBs), and the powers and responsibilities were given to an individual, the Director General. The director generals were appointed by government but couldn’t be sacked if their decisions proved politically inconvenient. The legal framework provided a fair balance between the interests of investors and customers. The regulator had duties to ensure that all reasonable demands for supply were being met, and also to ensure that the regulation of prices was established on a basis that the investor was expected to earn at least a normal rate of return equal to the cost of capital if the industry was managed cost-effectively.

The number of regulators reflected the distribution of duties across the United Kingdom, and the structure of the privatisations. Account also has to be taken of the fact that some thought that placing all the powers of regulation in the hands of a director general was an over-concentration

of power, and that it would be better to have a more traditional ‘Regulatory Authority’. This approach was adopted from 2000 onwards. The main dates the infrastructure industries were privatised and their regulatory bodies established follows, noting that in the first four privatisations the industries were not restructured before privatisation, but after that they were in order to better separate the natural monopoly elements in the supply chain from the potentially competitive. Major changes over time to the regulatory bodies are also shown.

Date	Industry	Regulator
1984	British Telecom	Office of Telecommunications (Oftel) became Office of Communications (Ofcom), 2003
1986	British Gas	Office of Gas Supply (Ofgas) merged into Office of Gas and Electricity Markets (Ofgem), 2000 with the Gas and Electricity Markets Regulatory Authority (GEMA)
1987	British Airports Authority (BAA)	Civil Aviation Authority (CAA) with Monopolies and Mergers Commission, now Competition Commission
1989	Water and Sewerage Authorities (E/W)	Office of Water Services (Ofwat) became the Water Services Regulatory Authority, 2005
1990 - 1996	UK Electricity Industry	Office of Electricity Regulation (Offer) merged into Office of Gas and Electricity Markets (Ofgem) with Office of Regulation of Electricity and Gas (NI) now NI Authority for Energy Regulation, 2003
1992 - 1997	British Rail	Office of the Rail Regulator (ORR) became the Office of Rail Regulation (ORR), 2004 with other bodies for the Train Operating Companies (TOCs) and the rolling stock companies (ROSCOs) being the Office of Passenger Rail Franchising (Opraf) merged into the Strategic Rail Authority (SRA), ended 2005 and the Office of the Arbiter for the London Underground public private partnerships (PPPs)

Over the years the structures of the utilities and network industries have changed to reflect the experience of independent regulation, and to facilitate more effective competition. BT has demerged to some extent to ensure that other providers of telecom services can use its natural monopoly over the ‘local loop’. British Gas demerged into its transmission and distribution pipeline business, and gas supply was bought by Centrica, still using the name British Gas in the UK. Railtrack was put into administration by the government in 2002 after a series of accidents which meant confidence was lost in it as a public service provider, and replaced by Network Rail, a not-for-dividend company. This model was also adopted for Welsh Water which effectively became bankrupt because it tried to run too many things other than its core water business. The traditional vertically integrated supplier in the different parts of the utility and network industries has often been replaced with horizontal integration across the various parts of the supply chain. Governments have continued to pursue the idea of privatisation of Royal Mail, and shares are to be offered for sale in 2013. The same political debate is felt in relation to the National Health Service (NHS) but here the emphasis really falls on facilitating more competition in supply.

3. Types of Competition

It is important to consider the types of competition that are available, otherwise it can be thought that ‘natural monopoly’ cannot be subject to some of the disciplines and incentives of a competitive market. Five types of competition are usefully identified.

* **Direct product or service competition**

This is competition ‘in’ the market, where providers can supply goods and services to customers who have a choice of which provider they would like to buy from. The usual incentives apply in matching supply to demand, and prices in general will be ‘cost-reflective’. Some barriers to direct competition might have to be regulated, such as where providers need access to a monopoly element in the supply chain. The provider of that monopoly element would be regulated to ensure that all users had ‘non-discriminatory’ rights of access to that common resource (hence ‘common carriage’).

* **Competition ‘for’ the market, rather than ‘in’ the market**

The classic example of this occurs in rail privatisation, where train operating companies bid for franchises to operate rail services in particular geographic regions of the national rail network for a certain number of years. The TOC which wins the franchise then becomes the monopoly provider for the period granted, but has to operate in compliance with the terms of the franchise they bid for, and that was conducted in a competitive market. To make this work well, the regulator has to have regard to the fact that many journeys will cross geographical areas, raising issues about providing comprehensive information, fair ticketing and fair redress (if a train is delayed, is it the fault of the TOC or Network Rail?).

* **Competition for procurement – contracting out**

Many organisations have a choice as to whether to employ their own staff to provide certain goods and services, or whether to invite outsiders to provide it for them. Local authorities have often had direct labour organisations, and it has been a part of government regulatory policy to require local authorities to test the costs and quality of providing their own services against those who would bid to compete, and to require the local authority to choose the provider which gives the best value for money. Contracting out is a very widespread form of competition.

* **Competition for corporate control**

Where an organisation has shareholders, then there is a strong incentive on the directors of the company to have regard to how well they run the company. Share capital tends to flow towards those who are best able to provide a good return, which means having regard to cost and quality over the longer term. An inefficient company would soon become the target of a hostile take-over bid, and its shareholders might be very willing to sell their shares and see a new management introduced. Given this incentive, the Government soon changed the restrictions on being able to take-over a privatised utility or network company.

The final type of competition, which has been very important in the regulation of the privatised utility and network companies is:

* **‘Surrogate’ competition**

This type of competition is where the regulator acts as a ‘surrogate’ for competition, in effect setting a price control regime which mimics what would happen in a competitive market. The importance of this is that it stops the abuse of monopoly power but gives the incentive to improve cost-effectiveness.

- Access pricing and the ‘efficient component pricing rule’

If the monopoly infrastructure supplier was to allow competition for supply where the new entrant would pay the infrastructure provider a fee for using their network, the determination of the appropriate ‘access’ price was an important part of the regulatory debate. If the public was to be better off from allowing competitive access, account has to be taken of both the

incremental costs to the new entrant and the ‘avoidable’ costs to the incumbent. The point of the efficient component pricing rule (ECPR) was to ensure that the cost of a new entrant using the incumbent’s infrastructure was not greater than the avoidable costs achieved by the incumbent from having a new entrant. It was not concerned with the separate question of whether the incumbent’s infrastructure business was efficient, only whether new entry was efficient. The main principle therefore was that access prices should be set to ensure that the new entrant’s incremental costs are less than or equal to the incumbent’s avoidable costs A^{CI} . In effect, to ensure a net gain resulting from a new entrant. This is efficient because the potential new entrant will only enter if the access price P^A , when added to the incremental costs of entry I^{CE} , total less than the incumbent’s current retail price P^{RI} . If the access price P^A is set equal to the current retail price less avoidable costs, that is, $P^{RI} - A^{CI}$, then the following condition applies $P^A + I^{CE} \leq P^{RI} - A^{CI}$, which is the same as saying the incremental costs of new entry must be less than or equal to the avoidable costs of having a new entrant $I^{CE} \leq A^{CI}$, the criterion to be met for efficient entry.

Setting the access price in this way gives the correct outcome based on the new entrant’s price incentives; any higher would be anti-competitive, and any lower would be inefficient. It might be the case that the incumbent’s supply business has avoidable costs which are lower than the new entrant’s incremental costs because it has significant economies of scope between the infrastructure and supply business in the supply chain. The ECPR ensures that the benefits of economies of scope are not lost through inefficient new entry competition.

4. Incentive Regulation – The role of the independent economic regulator

The purpose of the new independent economic regulators has been enshrined in statute, and developed overtime to reflect the experience of regulating the privatised industries, and to link them to broader policies of government. The primary objectives of consumer protection from the abuse of monopoly power, and consumer benefit from promoting more cost-effective industries, were reflected in the statutory duties to, first, promote competition (where appropriate) and, second, apply a system of ‘incentive’ regulation aimed at price control rather than profit control (the latter often being referred to as ‘rate of return’ regulation). Incentive regulation could be applied in the long run to those parts of the industry which were natural monopolies, and applied for a ‘transitional’ period to those parts which were potentially competitive, but required time for effective, sustainable competition to develop.

The economic regulator is subject to various statutory constraints, such as ensuring:

- * The regulated business can finance itself (based on the idea of controlling prices which would allow an efficient company to earn at least a normal rate of return equal to the cost of capital, given the risks the company faces);
- * Specified environmental and service standards can be achieved;
- * Monopoly networks are open to other providers to use based on non-discriminatory access tariffs;
- * Specified public and universal service obligations can be met;
- * Regard to special interest groups is paid (disabled, the elderly, low income families etc);
- * Duties towards promoting ‘sustainable development’;
- * Principles of ‘good regulation’ are applied by the regulator.

The list of constraints is to ensure that regulation is in the public interest, but one has been controversial, the one requiring the economic regulator to have regard to special interest groups. This arose because economic regulators could argue that they are not ‘agents of social justice’,

but surrogates for the discipline of a competitive market. Their job was to promote cost-effective utility and network services, and if the government found that certain groups of people were unable to afford such services, then this was the responsibility of government in terms of redistributing income. Tax and welfare support was seen as the responsibility of the elected government, not 'technocratic', appointed economic regulators.

Price control was chosen because it can achieve the same purpose as profit control to avoid the abuse of monopoly power (there would be no 'excess' profits if prices can be no higher than the long run cost of supply), but has the added advantage that it can reduce the long run cost of supply because there is an incentive on the management to 'out-perform' the regulators price controls. Inefficiency can be an abuse of monopoly power, reflected by one economist saying that "the best monopoly profits are a quiet life", something which nationalised industries are often accused of. Combining the two elements, control of excess profits and tackling underlying inefficiency, was therefore a telling reason for choosing price over profit control.

5. The 'Littlechild' model – Generally known as RPI –x

In February 1983 the Department of Industry published a report from Professor Stephen Littlechild, Professor of Commerce at Birmingham University, on 'The Regulation of British Telecommunications' Profitability. This is important because its recommendations set the regulatory framework for the forthcoming privatisation of BT, and the subsequent privatisations of utilities and network industries. The 'Local Tariff Reduction' scheme proposed (now known as RPI – x) was accepted by the government in place of the proposals which had been put forward by an interdepartmental working group and by a government adviser, Professor Walters. Although control of prices was an idea which had been used by government's in the past, it was in individual cases, and what made the Littlechild report such a seminal work was that it clearly highlighted the weaknesses of rate of return regulation compared with the incentive properties of RPI – x, but also because it could become the generic model for all subsequent privatisations. The report gave a ranking (from 1-5) of the comparative advantages of the various proposed regulatory schemes and the following shows the results for the working group's 'maximum rate of return' method compared with Littlechild's local tariff reduction scheme;

	Working group MRR	Local tariff reduction
Protection against monopoly	3	1
Efficiency and innovation	4	1
Burden of regulation	5	2
Promotion of competition	5	2
Proceeds and prospects	4	1

The comparison of proceeds and prospects related to the government's objective to maximise as far as possible the revenue from selling 51% of BT's shares.

In reading the Littlechild report it is important to be aware of the competitive context in which it was set. In effect, price control was a 'surrogate' for competition, and if competition could be rapidly introduced, then regulation of prices could be discontinued. The report has been criticised by some as not paying enough attention to what should happen if competition did not develop, either due to natural monopoly elements or government policy. It is true that the process of resetting x, and the period to which the price control should apply, were developed once the independent economic regulators were in place following privatisation, but this is not unreasonable given the Littlechild report was focused on combating the other proposals in a short period of time awarded by the government (some 12 weeks). Competition in

telecommunications did not develop as fast as some might have hoped, in part due to the government's policy of maintaining the duopoly between BT and Mercury, but when it comes to natural monopoly, Professor Littlechild was clear about the role of price control in his second report to the Department of the Environment on Economic Regulation of Privatised Water Authorities, published in January 1986.

Two important conclusions in the report are that "First, the water industry is largely a natural monopoly. Second, privatisation will involve not one but ten water authorities." (1.1). The first meant that "Competition is not an alternative to regulation of water authorities, but it is an important complement which can facilitate the regulator's task and provide added protection to customers" (1.6), the second that 'yardstick comparisons would be possible. In addition, "Regulation must encompass not only prices but also levels of service. The two must be considered simultaneously." (1.2). In terms of competition, there would be opportunities in "the product market for commercial services, as well as the provision of 'core' services on the borders between water authorities, and in 'franchising' or sub-contracting." (1.6). The report stresses the importance of allowing competition in the capital market, saying "The possibility of takeover ensures that each authority is under pressure to run its entire business efficiently and innovatively". (1.9). For this reason "The artificial prevention of takeover (eg, by a 'golden share') would shield inefficient management and remove the main protection for shareholders." (1.8).

Finally, whereas in Littlechild's first report "RPI – x was initially adopted as an interim measure for regulating BT, pending the development of more effective competition." (10.1), in regulating privatised water authorities "A permanent RPI – x system must provide for periodic revisions of x to prevent prices and costs getting too far out of line." (10.23). In choosing the price control period, the report concludes "Leaving it until sufficient pressure develops introduces political considerations and involves more work on monitoring. It would be preferable to have a periodic review at fixed intervals. A shorter interval reduces the risk of x lying outside the 'feasible range' but involves a greater burden on the regulator and may reduce incentives to efficiency. An interval of 5 years seems about the minimum time necessary to protect incentives; 10 years is about the maximum that shareholders, customers and government would find acceptable." (10.19). Five years has proved to be the period considered most suitable in practice.

The need for more permanent regulation through RPI – x has brought into perspective the relationship between price control and rates of return. In each control period there is no direct control of profit, which is shown in the rate of return achieved, but as each control period passes, using information revealed in the previous price control periods, there will tend to be a convergence between the rate of return achieved and the cost of capital. This has led some to say that rate of return regulation is 'ex ante' whilst RPI – x price control is rate of return regulation 'ex post', that is deferred. The following sections examine setting a price control in terms of annual accounted costs, the cash flows over the price control period (the truncated cash flow model), the profiling of prices over the control period and the relationship between the present value of costs and revenues over the control period, and comparing forecasts with outturns.

Setting a price control – annual accounted costs

To set a price control the regulator has to forecast the total costs which a regulated company would have to bear to provide the public service at the required quality of service. Given the three costs of business, it would be necessary to forecast the annual operating expenses (Opex),

the annual depreciation charges (on Capex), and the finance costs, the opening net book value (NBVo) of assets each year on which the cost of capital r^* would be paid in either interest on loans or dividends to shareholders. Added together, it would give the total costs (TC) for each year. Given these three expenses, forecast for each year (1, 2.....) of the price control period, the results can be discounted (df1, df2,...) to give the present values (PVC). The format would therefore be, for a five year control period:

<u>Year</u>	<u>1</u>	<u>2</u>	<u>5</u>	
	Opex1.df1	Opex2.df2	Opex5.df5	= Σ discounted Opex
	Deprec1.df1	Deprec2.df2	Deprec5.df5	= Σ discounted Deprec.
	NBV1o. r^* .df1	NBV2o. r^* .df2	NBV5o. r^* .df5	= Σ discounted finance
	Σ TC1.df1	Σ TC2.df2	Σ TC5.df5	= Σ disc.TCs = PVC

The opening book value each year is calculated from forecasting both the depreciation charges and the capital expenditure each year, given the closing book value each year equals the opening book value plus capital expenditure in the year minus depreciation charges for the year, noting too that the closing NBV equals the opening NBV for the next year:

<u>Year</u>	<u>1</u>	<u>2</u>	<u>5</u>
	NBV1o	NBV2o	NBV5o
	+ Capex1	+ Capex2.....	+ Capex5
	- Deprec1	- Deprec2	- Deprec5
	= NBV1c	= NBV2c	= NBV5c

and an illustrative worked example could show:

<u>Year</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	
NBV _o	100	110	119	133	152	
+ Capex	20	20	26	32	108	Σ 206
- Depreciation	10	11	12	13	14	Σ 60
= NBV _c	110	119	133	152	246	

So that an illustrative worked example of the undiscounted costs could show, where the cost of capital (r^*) is 10%:

<u>Year</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
Opex	20	19	18	17	16
Depreciation	10	11	12	13	14
NBV _o x r^* (10%)	10	11	11.9	13.3	15.2
Total costs	40	41	41.9	43.3	45.2

This 'annual accounts' method provides the forecast information needed to set the allowed revenue required for each year of the price control period, and to set the price control for each year, given the forecast quantity (Q^F) to be supplied at each price. It is also useful in that the forecast economic profit for each year can be related to the cost of capital r^* . Given revenue is set equal to total cost, then Revenue minus Opex minus Depreciation equals Operating Profit and the remaining revenue from above is NBVo times r^* , so that operating profit divided by capital employed (the return on capital employed) equals r^* . An illustrative worked example drawing on the above cost figures would show, where revenue a) = b) + c) + d), the three costs of business:

Year	1	2	3	4	5
a) Revenue (P.Q ^F)	40	41	41.9	43.3	45.2
b) less Opex	20	19	18	17	16
c) less depreciation	10	11	12	13	14
equals Operating profit	10	11	11.9	13.3	15.2
d) less NBVo x r*	10	11	11.9	13.3	15.2
equals economic profit	0	0	0	0	0
ROCE	10%	10%	10%	10%	10%

The economic profit is shown as zero because the regulator has set the price control such that PVC^F equals PVR^F so that $NPV^F = 0$. The return on capital employed (ROCE) would however be $\text{Operating profit} \div RBVo = r^*$.

The 'truncated' cash flow model

An alternative approach is to focus on the control period as a whole. Having chosen a five year control period, the regulator has to take account of the costs which the company is forecast to bear in that five year (ie, truncated) period. In financial terms, this would include the assets which the company has in place at the beginning of the control period, the cash flows on operating and capital expenditure in the five years, and the assets which would remain in place at the end of the period, assets which would not be a charge to the current control period. The asset values reflect what would be needed to buy the assets at the beginning of the period and sell them at the end of the period, hence it can be referred to as a truncated cash flow model. In accounting terms, the assets can be split into gross acquisition cost less provision for depreciation equals net book value (NBV). The NBV at the end of the control period would be the opening NBV plus the sum of the capital expenditure less the sum of depreciation chargeable, that is, $NBV_c = NBV_o + \sum \text{Capex} - \sum \text{depreciation charges}$.

This alternative approach, the 'truncated cash flow model', provides the same overall results as the annual accounts method, and brings in the capital expenditure for each year, whilst focusing on the total price control period. The format of this would be:

$$PVC(1-5) = NBV1_o + \sum \text{discounted Opex}(1-5) + \sum \text{discounted Capex}(1-5) - NBV5_c.df5$$

The relationship between the two approaches can be shown in the following table for a five year price control and year 1 for the annual accounts:

Over control period (1-5)	By annual instalments (1)	Profit and loss a/c (1)
+ NBVo1	+ Opex1	+ Revenue1
+ $\sum \text{Opex}(1-5)$	+ Depreciation1	- Opex1
+ $\sum \text{Capex}(1-5)$	+ NBVo1 x r*	- Depreciation1
- NBVc5	= Annual total cost 1	= Operating profit 1
= Total accounted period cost		

The truncated cash flow model, calculated a year at a time, would give the same results as the annual accounts method for the total discounted costs each year. Take year 1 for example:

$$PVC1 = NBV1_o + Opex1.df1 + Capex1.df1 - NBV1_c.df1$$

and because $NBV1_c = NBV1_o + Capex1 - \text{Deprec.1}$, then by substitution Capex 1 will cancel out and by rearrangement:

$$PVC1 = Opex1.df1 + \text{Deprec.1}.df1 + NBV1_o.(1 - df1)$$

the three costs of business used in the annual accounts method, because $NBV1o.(1 - df1) = NBV1o.r^*.df1$, given the discount factors are based on the cost of capital r^* .

Differing proportions of the three costs of capital

The only difference would occur when comparing the proportions which the three costs of business make up of the total discounted costs if the truncated cash flow method was not applied one year at a time. The total present value of total costs would be the same, but the proportional allocation of those costs would differ. Take a two year control period for example. By the same process of substitution and rearrangement the result would be:

$$\begin{aligned} \text{Opex1}.df1 + \text{Opex2}.df2 &= \sum \text{discounted Opex} \\ \text{Deprec1}.df2 + \text{Deprec2}.df2 &= (\text{Deprec1} + \text{Deprec2}).df2 = \sum \text{discounted Deprec.} \\ NBV1o(1 - df2) + \text{Capex1}(df1 - df2) &= \underline{\sum \text{discounted Finance}} \\ PVC1+2 &= \sum \text{discounted TCs1+2} \end{aligned}$$

and the difference occurs because the sum of the discounted depreciation falls compared with the annual accounts method (the depreciation charges are all discounted by $df2$ rather than the discount factor for the year in which the charge is made), and the sum of the discounted finance cost rises by an equal amount.

This can be demonstrated using an illustrative worked example for a two year price control period where $r^* = 10\%$; opex is 16 then 14, Capex is 40 then 60; depreciation is 20 then 20; and the opening net book value is 100. The discount factors, to four decimal places, would be 0.9091 and 0.8265 for years 1 and 2.

Annual accounting method (2 year price control)

Opex	$\text{Opex1}.df1 + \text{Opex2}.df2$	$= 16 \times 0.9091 + 14 \times 0.8265$	= 26.1166
Deprec.	$\text{Deprec1}.df1 + \text{Deprec2}.df2$	$= 20 \times 0.9091 + 20 \times 0.8265$	= 34.712
Finance	$NBV1o.r^*.df1 + NBV2o.r^*.df2$	$= 100 \times 0.1 \times 0.9091 + 120 \times 0.1 \times 0.8265$	= <u>19.009</u>
	Total discounted costs (PVC1+2)		= 79.8376

which compares with:

Truncated cash flow method (2 year price control)

Opex	$\text{Opex1}.df1 + \text{Opex2}.df2$	$= \text{as above}$	= 26.1166
Deprec.	$(\text{Deprec1} + \text{Deprec2}).df2$	$= (20 + 20) \times 0.8265$	= 33.06
Fin.	$NBV1o(1 - df2) + \text{Capex1}(df1 - df2)$	$= 100(1 - 0.8265) + 40(0.9091 - 0.8265)$	= <u>20.654</u>
	Total discounted costs (PVC1+2)		= 79.8306

In the truncated cash flow model the total discounted depreciation costs have fallen by 1.652, but the finance costs have risen by 1.645. The difference between the two is 0.007, which is a 'rounding' error resulting from using approximate discount factors based on four decimal places. The 0.007 is also the difference between the total discounted costs.

RPI - x

The term $RPI - x$ simply means that the controlled price each year has to relate to the price last year times the RPI index for the current year over the last year minus $x\%$. Given the closing price of the last control period (P_c), the regulator will determine the $x\%$ for each year to be applied, for example, -3% , -3% , -2% , -2% , -2% for a five year control period. Given the forecasts might require real increases in prices to fund increased investment, then it might be $RPI + k$. Separating the price increases to compensate for the fall in the value of money though inflation from the real terms cuts (the $x\%$ s) or increases (the $k\%$ s) is important as it enables the regulator to prepare the forecasts for the next control period in terms of current prices. This means that prices may appear to go up, but this is only in nominal terms because of inflation and

the $x\%$ or the $k\%$ reflects the real change in prices. Each years price change would therefore be last year's price cap $\cdot (1 + \Delta RPI) - x\%$ or $+ k\%$. Assuming inflation was 10%, $x = -3\%$ and last years price cap was 12, then the next years price cap would be $12 \times 1.1 \times .97 = 11.804$.

Sometimes, to maintain the distinction between efficiency improvements and investment increases, a regulator has published the price controls as $RPI - x + Q = RPI + k$. Given three control periods then we could have:

Control period 1	Control period 2	Control period 3
2007 - 2011	2012 - 2016	2017 – 2021
@ 2006 prices	@2011 prices	@2016 prices

The impact of inflation is important for determining the appropriate accounted costs which apply to each price control period, given that many costs arise from the capital assets which have a life longer than the current price control period. The acquisition costs for an asset many years ago might differ greatly from its replacement cost now, and this affects the depreciation charges appropriate for each price control period. If depreciation charges are based purely on acquisition costs (historical cost accounting), then are current customers paying their fair share of the depreciation in use of an asset which will cost far more to replace? To ensure that customers pay their fair share, it is normally required for regulatory accounts to be prepared in current cost terms. This accords with the accounting principle of Financial Capital Maintenance (FCM), which is intended to maintain the value of the business in real terms by requiring the profit and loss account to have its depreciation expense charges based on the replacement costs of the fixed assets. Clearly this is an important concept for the utilities and network industries, so reliant on their infrastructure.

However, there are two sides to regulating these industries: the customers and the shareholders. It is fair to say that customers should pay the current cost of using the fixed assets, but would it be fair to say that investors (the shareholders) should receive a return on assets valued at more than the investors paid for them (the acquisition costs)? This is important for regulatory accounts, where a distinction is made between the net book value (NBV) of the assets and the regulatory net book value (RBV) which is used to determine the price controls. The accounts are allowed to be revalued to current replacement costs, but for those assets which were in existence at privatisation, if they were acquired for less than their current replacement cost at privatisation, then the NBV is reduced by that discount in absolute terms. This ensures that shareholders do not receive a 'windfall' gain simply from allowing the assets to be revalued to current replacement costs. The regulatory accounts would therefore include:

	Gross current cost of assets	x
<i>less</i>	provision for depreciation	(x)
<i>equals</i>	current cost net book value	x
<i>less</i>	discount at privatisation	(x)
<i>equals</i>	regulatory book value (RBV)	x

The price control period and the 'regulatory bargain'

To give the incentive to improve efficiency, sufficient time must be given to the regulated company to achieve out-performance and receive some additional profit as their reward. The price control has therefore to be set for a period, based on 'forecasts' of the cost of supply, covering the three costs of business – operating costs, capital consumption, and the cost of capital finance. The question arises, how long should the control period be? Regulators have tended to set their control periods for five years, being long enough to allow a sufficient incentive for out-performance, but not so long that if the forecasts prove unreliable, it would be

politically untenable to continue, either because the regulated company was being driven into bankruptcy, or because much too large excess profits were being earned for too long.

The control period chosen represents a ‘regulatory bargain’ between shareholders and customers. The economic profits earned from out-performance for the control period are in exchange for ‘revealing’ the improved efficiency (reduction in the long run cost of supply) which can then be used by the regulator in setting the price controls for the next period, thereby passing on the economic profits to customers. This regulatory bargain can be misunderstood, because the additional profits come first. The answer, however, has to be that without the incentive to out-perform, there would be no reduction in the long run cost of supply to pass on to customers. This has a regulatory consequence, which is that regulation must be ‘forward looking’, so that we should not expect regulators to be backward looking for the purpose of clawing back additional profits made from out-performance. Otherwise the incentive is lost and the mutual benefit for shareholders and customers is lost.

The question might be asked, ‘how equitable is the balance between shareholders and customers in what is received from out-performance?’ The following calculation is an example.

Supposing, having price controls set on forecast costs for five years, the regulated company makes a cost saving of £1 per year compared with the forecast. If the cost of capital was 7%, then the present value of £1 receivable for five years would be £4.1. Assuming this revealed saving is then used by the regulator for all subsequent price control periods, then the benefit to customers is the present value of a £1 saving in the long run less the £4.1 taken by shareholders. This is $(1 \div 0.07) - 4.1$ which is $14.29 - 4.1 = £10.19$. The proportion going to shareholders is therefore $4.1 \div 14.29 = 28.7\%$, and the proportion going to customers would be 71.3%. Whilst it is a simple example, it shows why price control can be considered to be an equitable bargain.

Variance analysis of forecasts to actuals

Having calculated the present value of forecast costs (PVC^F), given the cost of capital r^* , the price control can be set such that present value of forecast revenue (PVR^F) is equal to it, that is, $PVR^F = PVC^F$, and this means that the forecast net present value (NPV^F) is zero. At the end of the control period the forecasts can be converted to actuals (outcomes), and the actual net present value for the period calculated. The difference between the forecasts and the actuals will be a measure of the economic profit or loss. These variances are important in determining the next period’s price control. The relationship is usefully shown as:

$$\begin{aligned} PVR^F &= PVC^F \\ \frac{PVR^A}{\Delta PVR} &\leq \text{or} \geq \frac{PVC^A}{\Delta PVC} \\ \Delta PVR + \Delta PVC &= \text{Total variance} = \text{economic profit or loss} \end{aligned}$$

Variance analysis is important because the regulator needs to learn the lessons of the past control period in order to inform setting the price controls in the next. In some circumstances it would be legitimate for the regulator to take a ‘backward’ look and claw back some of the economic profit earned. This would be where the profits were earned ‘illegitimately’. One case would be where the price cap had been set to allow the regulated company to improve the level of maintenance or invest in new capital assets, but the company had not incurred those expenditures whilst at the same time taking the revenue allowed by the price cap. However, this should only be clawed back where that non-expenditure had really compromised quality or the security of supply in the current control period. Otherwise it could stand in the way of the incentive to get the most out of the existing assets, for example, by extending their working life. If this is a good engineering decision, then the additional profit is legitimate and the customers benefit in the longer run through lower capital costs. Regulators might also not allow the cost of ‘uneconomic’ purchases of fixed assets to be recovered in subsequent control periods.

Clearly, deciding on such things may not be easy, and would often result in disputes between the regulator and the regulated. In the end it is having transparent and accountable processes for determining the outcome which makes it manageable. One problem of having a fixed period for the price control is that the incentive to improve efficiency in the last year of the control period can be less than in the first year. For this reason, some regulators allow some 'carry-over' of out-performance into the next period in order to maintain the marginal incentives to out-perform throughout the current period and enhance the incentive to out-perform overall.

The basic financial 'template'

Before showing an illustrative example of the basic financial template two further aspects have to be added in. First, the capital expenditure is divided into that concerned with 'replacing' existing assets, and that concerned with 'enhancement'. A new European directive might require completely new treatment plants for disposing of used water into rivers and seas, and it is useful to keep this separate from pure replacement of existing assets. In fact, it is such changes in government policy which give reason for regulators to reopen the current price control period and set new price caps based on meeting the obligations arising from policy changes.

Secondly, a distinction between the net book value (NBV) and the regulatory book value (RBV) is introduced, given any discounts to investors at the time of privatisation. Subtracting the discount from the NBV means the PVC is reduced and customers only pay for what investors actually paid for the assets. The discount has to be subtracted from closing NBV as well, and the amount will depend on whether the regulator is carrying through the discount as a constant into each control period, or whether it is being written off through depreciation charges. The present value of costs added over all the control periods remains the same under either method, but the distribution of costs, one period with another will change.

In calculating the net book value any written off (w/o) assets in the control period cancel each other out in terms of gross acquisition cost and provision for depreciation, so that the NBV has only to be concerned with capex and depreciation charges for the period. The regulatory book value for the closing of the price control period in the undiscounted cost series is shown as a negative number (often shown in brackets), being a notional receipt as current customers are not liable for these costs in the current period. The discount factor for the opening of year 1 would be equal to 1.

The Basic Financial Template							
Year (opening/closing)	1o	1c	2c	3c	4c	5c	5c
OPEX		20	19	18	17	16	
CAPEX							
Replacement		0	0	6	12	108	
Enhancement		20	20	20	20	0	
Gross acquisition cost	200	+ \sum Capex 206 - w/o assets 126					= 280
less depreciation provision	(100)	+ \sum deprec. 60 - w/o assets 126					= (34)
<i>equals</i> net book value	100	+ \sum Capex 206 - \sum deprec. 60					= 246
less discount@privatisation	(50)	carried through unless w/o					(50)
<i>equals</i> regulatory book value	50						196
Undiscounted cost series	50	40	39	44	49	124	(196)
<i>times</i> discount factor (7%)	1	0.935	0.873	0.816	0.763	0.713	0.713
= Present value of costs	50	37.4	34	35.9	37.4	88.4	(139.7)
and the sum of the discounted cost series \sum PVC = 143.4							

Profiling the price controls over the control period

To fill in the figures in the financial template would usually take the regulator's office at least two years. A business plan has to be obtained from the regulated company, and the regulator has to carry out various comparative studies to see how this regulated business compares with others, in this country and abroad. The methodology to be used has to be decided on, and consulted upon, and the cost of capital appropriate has to be agreed. The regulator can build in expected efficiency savings into the forecasts (eg, operating expenditure should reduce costs by so many % per year). Once a final decision has been made, the proposal is submitted to the regulated company to accept or reject as the basis for the next price control period. If accepted, it proceeds, if not, an appeal goes to the Competition Commission which reports whether it agrees or disagrees with the regulator. The political economy of price control regulation has been tested many times by this method, as it has too by reports from the National Audit Office following inquiries by the select committees of parliament.

Having established the forecast present value of costs (PVC^F) for the control period, the regulator has to decide how the equivalent forecast present value of revenue (PVR^F) should be spread over each of the years. Some regulators adopt a 'revenue' cap, which might be appropriate when the fixed costs of the business are very high, and the variable costs very low, because if the actual quantity supplied is very different from the forecast quantity, a price control based on average revenue might lead to very large excess profits. It is for this reason that 'two-part' tariffs often apply in utilities and network industries (ie, a fixed annual charge plus a much lower rate per unit supplied, as in telephone charges). However, a price rather than revenue cap does give the incentive to the regulated business to supply what customers want.

One natural way to determine the profile of the price control is to link it to the forecast annual accounts which arise from the truncated cash flow model. Instead of being concerned only with the opening and closing book values at the start and the end of the control period, it is done for each year. This means a table is required for the net book values and the annual depreciation charges. The great advantage of this approach is that if the allowed revenue for each year was set equal to the PVCs for each year, then the accounts would show a forecast rate of return equal to the cost of capital. As shown, if the total annual costs are discounted at r^* then they equal in total the PVCs from the truncated cash flow model. If the allowed revenue for each year is set equal to the total costs for each year, then the accounts would show a normal rate of return equal to r^* . This is a great advantage when looking at the actual accounts in due course because you have an understandable comparator for their actual rate of return. Given the forecast revenue and quantities, the price cap for each year would be forecast revenue \div forecast quantity. The first year price cap (P_1) might be a substantial reduction on the closing price from the last control period (P_c) due to the 'passing on' of out-performance in the last control period, but if expenditure needs to be increased dramatically to improve quality then the traditional $RPI - x$ may become $RPI + k$, as it did in the water industry.

The flexibility in profiling the price control over a five year period can be shown from three separate price paths drawn from a worked example with the same present value of costs of 2000 and the closing price of the last control period equal to 5.

	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
5% reduction from P_0	4.75	4.75	4.75	4.75	4.75
2% reduction per annum	4.9	4.8	4.71	4.61	4.52
8% reduction from P_0 and then 1% increase thereafter	4.6	4.65	4.69	4.74	4.79

Which accounting model?

The regulatory accounting model is chosen to reflect the model used to set the price controls. Profiling the revenue over the price control period is one example of that, but there is always the question of profiling across various price control periods. Three broad models have been used, each of which has a different profile of revenues across a series of price controls, although all three are consistent with the underlying truncated cash flow model and the constraint of ensuring that the present value of allowed revenue equals the present value of forecast costs. The three models are:

- * The standard financial capital maintenance model: This is based on allowing the acquisition costs (ie, what investors paid for their assets) to be indexed to cover inflation, and with the depreciation charges based on the indexed acquisition costs. This means that current consumers would be paying prices based on acquisition costs which might have been heavily discounted at the time of privatisation compared with their replacement costs. Over time, as assets acquired at privatisation are replaced, the depreciation charges will rise to more reflect current replacement cost.

- * The regulatory hybrid model: This was developed to overcome the fact that current consumers were not facing the current replacement cost of using the assets in the standard model (a form of historical cost) and that this distorted their purchasing decisions. By revaluing assets to their replacement cost, and basing depreciation charges on those values, price controls could reflect their real economic cost, but the revaluation alone would mean that investors were receiving a return on assets which they had not paid for. To overcome this, the regulatory book value was reduced (abated) in price controls by the percentage discount at privatisation, so that the return was based on the cost of capital applied to the acquisition costs, whilst the depreciation charges were based on their replacement cost, and could be retained in perpetuity. The difference in the revenue profile between the two models would be that the standard model had lower prices for current consumers, and higher prices for future consumers, whereas the hybrid model would be higher prices now but lower than the higher, long run prices in the standard model.

- * The renewals provision model: This was developed to avoid depreciation charges, based on the idea of the infrastructure assets being a perpetual asset, as long as there was a maintenance charge in the price control to maintain the serviceability of the infrastructure assets in the long run. The renewals provision method provides a return in perpetuity on the indexed historical acquisition cost at privatisation; all other costs are met through the renewals charge. The hybrid and the renewals models have on occasion been combined, such as in the water industry where underground assets were treated as 'non-depreciable' and overground assets which were depreciated in the normal way.

6. Economic history 1980 – 2013

The economic history of independent economic regulation began in 1979 when the new Conservative government replaced the Labour administration. A Labour party had established the nationalisation of utilities and network industries following the end of the second world war in 1945, and that model had remained in place until 1979. Whilst the privatisations which followed established a major change to the role of the 'regulatory' state, and developed the model of public services privately supplied, the way in which it developed over time was a combination of experience and changes of government. For this reason it is best to divide the period into three distinctive periods: 1980-1997; 1997-2005; and 2005-2013, noting that 1997 was the change from a Conservative Government (essentially for privatisation) to New Labour (essentially against privatisation). This section will look at it from a chronological perspective, and citing material drawn from reports prepared by government departments, independent economic regulators, agencies and auditors, including select committees of the houses of

parliament. Although this economic history focuses on the utilities and network industries, it is important not to overlook other cases affecting regulation, such as the role of the economic arbiter in overseeing contracts for London Underground infrastructure and services, it not being part of Network Rail's responsibilities. There is also the increasing debate on, and regulation of banks, following the financial crash in 2008. There are numerous sources for this in particular. More academic commentary will be left to section 7, the overview and bibliography. Whilst it is convenient to divide the section into three distinct phases it is also important to recognise the continuity of independent economic regulation of privatised utilities over the whole period. The most important source for this period is **The Official History of Privatisation, Volumes 1 and II, by David Parker**, and published by Routledge in the Government Official History Series. The official series gives the authors access to all relevant official documents, and so can produce an important insight into how policies were discussed and implemented over time.

The most notable continuity was that when New Labour gained power in 1997, after eighteen years of Conservative governments; there was no action to renationalise, even though "throughout the 1980s opinion polls in Britain suggested that a large section of the public remained sceptical of the case for privatisation." (vol 1 p448). "After 1987, Labour Party and Trades Union Congress annual conferences made continuing but essentially ritualistic attacks on privatisation. The prospect that the privatisations would be reversed was relegated to the political margins." (p449). The Labour leader Neil Kinnock in 1987 said "if this movement pretends, for instance, that a few million people owning a few shares each will not make a difference to their perception of their economic welfare then this movement will be fooling itself....and the result of it is that our policies are going to have to take account of that reality.". "The Party's criticism of privatisation now focused less on the principle of privatisation and more on the way that the nationalised industries were being sold and on the large management salaries that resulted." (p449).

1980-1997

Before 1979 the financial and economic performance of nationalised industries had been open to question and much criticism. The new Conservative government's immediate policy had been to reduce the amount of subsidy paid to them and to control their borrowing requirement for capital spending by introducing 'external financing limits' (EFLs). In addition, in order to introduce competition as a means of promoting efficiency, and to 'roll back the frontiers of the state' the main thrust of government policy was to begin a process of privatising as many of the nationalised industries as possible. At first this was more concerned with organisations other than the utilities and network industries, but these soon came into play given their size, and privatisation also gave the opportunity to tackle issues of union power. Each privatisation of the utilities and network industries was organised to achieve as wide a share ownership by the British public as possible; hopefully as one minister said "to achieve democratic capitalism".

Between 1981 and 1984 the legal monopoly of British Telecommunications (BT) was dismantled in preparation for sale, and the first independent economic regulator, the Office of Telecommunications (OfTel) was formed in August 1984 to oversee the future regulatory framework. However, in practice, competition and the regulatory framework were influenced by the government's wish to maximise the proceeds from sale. The more competition, and the tighter the regulatory regime, the lower would be the proceeds from sale. This was one reason why the government maintained a duopoly in telecoms until at least 1990, and when the British Gas Corporation (BGC) was privatised in 1986, it was privatised as a single entity, not broken up to facilitate competition. One of government's aims, to keep an involvement in the privatised industries, was reflected in the Nationalised Industries' Legislation Consultation

Proposals issued by the Treasury in December 1984, which included a government power to hire and fire board members, but it was so opposed by the industries it wished to privatise, that the government was forced to drop the proposals in 1985. Before more detailed coverage of BT, British Gas, the BAA, the water industry, the electricity industry, and British Rail, the following refers to privatisation as a whole and the introduction of other regulations concerning quality of service. The regulators mentioned are those in place at the time, even though many have had their names and responsibilities changed over time or been merged with others. Examples include Oftel becoming Ofcom, and Offer and Ofgas becoming Ofgem (see Part 1V, section 2). Such things continue, such as ORR becoming the Office of Rail and Roads.

- the privatisation programme

Published in 1995 was Her Majesty's Treasury Guide to the UK Privatisation Programme. Covering most of the first period, 1980-1997, it provides a systematic review of what happened and reasons behind it. The appendices are detailed, including the major sales so far (A), the nationalised industries control framework (B), proceeds of privatisations and share sales (D), outline of typical steps in a privatisation (F), and an article (K) by the financial secretary to the Treasury, John Moore, and published in the Harvard Business Review (Jan-Feb 1992), and a speech (L) given by the financial secretary Sir George Young to the World Privatisation Conference in March 1995. Selected material drawn from the guide gives a clear view of the intentions and progress in government policy, starting with the typical steps in a privatisation:

Typical steps – Stage 1, feasibility study, report and ministerial decision; Stage 2, select advisers, prepare the business, prepare the legislation and consider regulation/deregulation; Stage 3, consider the balance sheet; Stage 4, select advisers for sale, choose market slot, produce prospectus and build the image by advertising, decide price; Stage 5, sell and transfer ownership.

The aims of privatisation – “The promotion of wider and deeper share ownership – both among employees and the general public – is part of the government's policy of extending the ownership of wealth more widely in the economy, giving people a direct stake in the success of British industry, and removing the old distinctions between ‘owners’ and ‘workers’. The privatisation programme has contributed greatly to an increase in the number of private investors in the UK from 3 million in 1979 to about 10 million in 1994. About 90% of eligible employees typically become shareholders in their companies on privatisation.” (9).

The mechanics of privatisation – “The preparations for privatisation are complex, and each case raises its own problems. A key issue is whether the aims of the privatisation would be achieved more effectively by restructuring an entity before sale. In the case of some major industries, significant restructuring may be directed to developing the potential for competition after the sale. For example, the former nationalised electricity industry in England and Wales was split into three generating companies, twelve regional electricity companies, and an entity owning the national grid. All bar the nuclear generating company were privatised in 1990-91. This structure allowed for competition in electricity generation and the distribution of electricity to major users.” (16). “Looking to future sales, as a prelude to the privatisation of the railways, track and train operations have been separated and a new government owned track authority, Railtrack, has been established. Operation of passenger transport services will be transferred over time to the private sector by franchising; BR's freight and parcels operations will be sold outright; and the private sector will, in addition, be given the opportunity to introduce new services, for both passenger and freight, through open access. Railtrack itself will be sold by flotation during the lifetime of the present parliament. For the privatisation of the nuclear

electricity companies, scheduled to take place in 1996, the old Magnox stations will remain in the public sector while the newer stations (further from decommissioning) will be sold.” (17).

Sale options – “Once any restructuring has been undertaken, the main options for the sale process itself are: a trade sale, where a company is sold to a single firm or consortium, including the option of a management/employee buyout; a public flotation on the stock exchange; or, rare in the UK government’s privatisation programme, a placing with a group of investors.” (18). “Sales of smaller entities have generally been targeted at trade purchasers, but with encouragement for management/employee buyouts.” (19). “Privatisation by flotation has been the option adopted for larger businesses where there was expected to be a wide range of institutional and/or public demand. As shown in major sales so far (appendix A), the principal methods of sale have been a fixed price offer of shares; a tender offer at a ‘striking’ price; or, in a number of cases, a combination of the two.” (20). “Government policy has been to give priority to the small investor; and to this end it has pioneered the development of major retail offer structures from the privatisation of BT in 1984, then the biggest public offer in the world, to the 1993 sale of virtually all the government’s residual holding in BT, which further developed the ‘Share Shops’ initiative (the appointment of eight financial service providers, with a high street presence, to provide cheap and accessible buying and selling arrangements for a range of shares) first ventured in the sale of BT shares in 1991”. (21).

The special share – “In a number of privatisations there has been a clear need to protect a business from unwelcome takeover, for example, on national security grounds, or, as a temporary measure, to provide an opportunity for management to adjust to the private sector. To this end the government retains, or has held, special (‘golden’) shares in a number of privatised companies. The special share requires that certain provisions in the Articles of Association of the company may not be changed without the specific consent of the special shareholder. The provisions typically include, for example, a prohibition on any one person, or groups of persons acting in consort, controlling more than 15% of the equity of the company”. (25). In 1995 the government’s holding of non-time limited special shares was listed in Appendix G as: Cable&Wireless; Sealink Stena Line; BT; British Aerospace; VSEL Consortium; British Gas; Rolls Royce; BAA; National Grid Company; National Grid Holding; National Power; PowerGen; Scottish Power; Scottish Hydro-Electric; Northern Ireland Electricity; Belfast International Airport. Two had an expiry date of 31 March 1997, British Technology Group and British Technology Group Int.

Regulation of privatised utilities – “The desire to distance the UK’s privatised utility industries from political control and to separate ownership and regulatory functions led to the development of regulators who undertake their responsibilities independently of government. Each office is headed by a Director General appointed by the relevant Secretary of State for a fixed term, normally 5 years. Each regulator has wide-ranging powers and duties to promote the interests of consumers while ensuring investors earn adequate returns on capital.” (43). “The independent regulators established by 1995 were: Office of Telecommunications; Office of Gas Supply; Office of Water Services; Office of Electricity Regulation; and Office of the Rail Regulator.” (53). “The remaining major privatisations in hand in 1995 were British Rail and Nuclear Electricity.” (56).

Achievements of regulation – “The championing of the consumer by the regulators has placed the utilities in an environment that encourages a better service than was provided while in the public sector.” (51). An important source of information on the financial success of the privatisations are the reports prepared by the National Audit Office for the Public Accounts

Committee (PAC) of the House of Commons. Two reports concerned the sale of the second and third tranche of shares in British Telecommunications plc, one in 1993, the other in 1994. Their conclusions in relation to the second tranche were that “Having regard to the size and timing of the sale, the Treasury did all they could to maximise sale proceeds (12); Dispensing with underwriting is likely to have significantly increased net proceeds.” (by reducing costs) (16); By comparison with previous sales, the higher rates of share allocations to individual investors and the higher levels of share retention by them in the months following the sale indicate that, in spite of the disposals evidenced by the share shops, the Treasury met with a broad measure of success in widening and deepening share ownership.” (21). The innovations introduced by the Treasury in this sale enabled them to address a number of problems which had beset previous flotations, in particular the pricing of the offer.” (22). In terms of the third tranche “The National Audit Office concluded that the Treasury successfully achieved their sale objectives, in particular, maximising proceeds and minimising costs; widening and deepening share ownership; and ‘perception’ of success.” (9).

- The development of quality regulation

Following BTs privatisation in 1984, it soon became apparent that not including the direct regulation of quality of service was unsustainable, and by 1991 the government introduced the Competition and Service (utilities) Bill which would give new powers to regulators to set quality standards and to determine levels of compensation to be paid to customers if those standards were not met. This period was well covered in a paper by Fod Barnes published in the Consumer Policy Review in 1992 (vol 2, no 1) entitled Quality Regulation: the UK experience of regulating BT. The early years experience of a privatised BT seemed to prove his comment that “The disadvantage with price-cap regulation is that produces a powerful financial incentive to reduce quality levels, since savings made through quality reduction also show up as increased profits. Indeed, in some circumstances it may be easier to achieve cost reduction through quality reduction rather than through increased efficiency”. The decline in the reliability of the red public phone boxes soon brought the issue to public attention, and the fact that there was limited competition to BT services meant that quality levels fell within the domain of its monopoly power. The quality problem was exacerbated because “at the time of privatisation, BT had been publishing a range of quality measures in its ‘report to customers’ or its annual report, but, for the year 1984-5 (the first time the results would have been published by the privatised company), BT discontinued the publication of these indicators on grounds of commercial confidentiality”.

Oftel was able to respond to this and “in early 1986 the Director General published a consultation document on quality of service regulation, and in November his first report on the quality of the telephone service”. BT agreed to begin publishing its quality of service measures again in 1987. One reason was that “1987 had been a bad year for BT – a combination of a strike, a customer survey from the National Consumer Council showing that BT was considered to be the worst of the utilities, and a serious problem with pay phone reliability, put BT in a position where it was politically impossible not to be seen to be doing something about its quality of service” (and pay phone reliability was increased from 77% in 1987 to 95% in 1990). BTs Customer Service Guarantee Scheme was introduced in 1989, not based on a change to its licence, but based on a contract between BT and its customers in which BT took on some limited liability for the quality of its services, with compensation payments of £5 for each additional day over the standard set for the service, and up to £1000 (residential) and £5000 (business) for actual financial loss caused by a service delay beyond the allowed two days.

In general terms, the link between quality of service and the price control formulae was being established. The Director General could make it clear that reductions in quality would be seen as 'hidden price rises', and take appropriate regulatory action at the next price review, or even by reviewing the existing price formula. In 1992 the Competition and Service (utilities) bill became law, and the integration of quality and price regulation had been achieved. The National Audit Office published its report on The Office of Telecommunications: Licence compliance and consumer protection in 1993, stating in paragraph 2.13: "In 1992 the Director General reviewed the practicality of linking BT's pricing policy to service quality but concluded that a workable link did not exist, not least because of users' differing perceptions of service quality and value. He considered that improvements in quality were better secured through monitoring and publishing BT's performance statistics and using his new powers under the Competition and Service (utilities) Act. The Director General also gave BT a specific warning that if a significant deterioration in service quality were to occur, this would be regarded as a reason for looking again at the level of the price cap. Oftel accepts the principle that, equally, improvements in service quality are effectively a price reduction".

The development of quality regulations in telecommunications was a debate between Oftel and BT with a view to protecting customers from the abuse of monopoly power. The next most important quality case, however, was between the new regulator for the water industry and the government, in a sense to control the monopoly power of the government in the interests of water customers. The issue was that governments, when setting new quality standards, should take account of both the costs and benefits of those new standards, and the formation of the new 'independent' economic regulators gave the opportunity for the balance to be publicly debated, making the process more transparent and accountable. In this case the Director General of Water Services published in 1992 'The Cost of Quality – A Strategic Assessment of the Prospects for Future Water Bills'. In the foreword, it said:

"Improvements in the quality of drinking water and the water environment are beneficial. But they cost money – which many customers can ill-afford. Our evidence of customer views shows that while they are concerned with water quality and the environment, they are also concerned about bills. The issue is particularly acute when much of the cost is incurred in making high standards even higher – standards which, it is sometimes argued, are not always securely based on scientific evidence of the risks to health. Customers might ask whether the drive for higher quality should go quite so fast".

After further work on the forecast costs by Ofwat, the Director General published in 1993 'Paying for Quality – The Political Perspective'. In the Foreword it said:

"My purpose in preparing this report is to seek from the Secretaries of State for the Environment and for Wales a clear perspective on new obligations and their implementation, as a basis for setting revised price limits. If price limits are to be soundly based, I need clear guidance from them on the new obligations that companies are likely to face and the time scales for achieving them. When these obligations are established I have a legal duty to ensure that they can be financed. To do this I must set price limits which can accommodate the necessary expenditure". The foreword ended: "I do not believe that customers – business customers as well as domestic customers – will regard the increases in bills set out in this report as affordable".

Politicians could not now escape the responsibility for imposing higher water charges on customers, and the improved form of governance continued in future periodic price controls. For example, in June 1998, Ofwat published 'Setting the Quality Framework – An Open Letter to the Secretary of State for the Environment, Transport and the Regions and the Secretary of State for Wales', and in September 1998, the Department for the Environment, Transport and

the Regions published 'Raising the Quality – Guidance to the Director General of Water Services on the Environmental and Quality Objectives to be achieved by the Water Industry in England and Wales 2000-2005'.

- Price controls on BT 1984-1997 and development of competition

The price controls on BT over this period showed that 'x' meant what it implied, a reduction in allowed real prices for certain telecommunication services. The range of services falling into the 'basket' also increased as each price review took place. The following table sets out a summary:

The price cap and dates	Elements in the basket	% of BT turnover under price control
1984-1989 RPI – 3%	* exchange line rentals * local and national dialled call charges	48%
1989-1991 RPI – 4.5%	* exchange line rentals * direct call charges * operator assisted calls and directory enquiries service	54%
1991-1993 RPI – 6.25%	* exchange line rentals * direct call charges * operator assisted calls * international calls * low users scheme	69%
1993-1997 RPI – 7.5%	* exchange line rentals * direct call charges * operator assisted calls * international calls * connection charges	71%

One thing the table shows is that the hopes at the time of privatisation that competition to BT would develop quickly enough to make price controls unnecessary after a short time were unfounded. However, the pace changed with the ending of the duopoly on telecommunications services over fixed links within the UK between BT and Mercury. By 1991 the impact of Mercury's competition had been to reduce BT's share of the UK business market for telephone calls and provision of exchange lines to 92% and for international calls to 83%, and the potential for increased competition was demonstrated by the government receiving 23 applications for new licences by January 1992, so that the policy debate started on whether the duopoly on overseas calls should be ended. To facilitate competition Oftel issued a policy statement in June 1992 on separation and interconnection which concerned giving more information to competitors who were negotiating with BT for access to BT's monopoly networks. Separate profit and loss statements and balance sheets were to be made available for the different BT businesses by 1993/94 and include, for the first time, current cost accounts. Details on how BT allocated its costs would be published, as well as information on interconnection agreements concluded by BT. Charges to competitors for the use of the BT network would have to be equally applicable to BT's own call business, thereby controlling cross-subsidy or discriminatory pricing. Accounting separation was particularly important in relation to BT

because Oftel had confirmed in the 1993 third share sale that “it is not seeking a structural separation of BT.”

These developments affected regulation of price controls because competition is intended to make prices cost-reflective, and one issue was the appropriate cost for making a call compared with the cost of renting a line, given that the prices charged must be sufficient to recover both fixed and variable costs. Call charges can be based on a combination of the incremental cost of making a call plus a contribution towards meeting the fixed costs (which can be collected by a standing rental charge per year in a two part tariff), but should the contribution be calculated as an equal percentage addition to the cost of each call, or take account of the demand for different types of call to minimise the loss of consumer surplus in recovering fixed costs? In the 1992 price review Oftel accepted BT’s argument that rental and call prices should be rebalanced, but argued that because of the uncertainty on the basis for cost-allocation, a slower pace of change was justified. BT’s access charges could include an ‘access deficit contribution’ to take account of the fact that BT’s rental charges are price capped, which stops BT eliminating the current cross-subsidy of rental charges from calls. Without the deficit contribution BT would not be covering its network costs, and hence would be subsidising its competitors. However, Oftel has waived these deficit contributions in some of its determinations on negotiations between BT and competitors in order to encourage new entry.

By 1996, however, regulation of BT’s prices had begun to move towards deregulation because competition had increased significantly since privatisation in 1984, and because developments in technology were so high that any natural monopoly was being eroded. Oftel had reduced the percentage of BT’s turnover which is controlled by price caps from 65% to 25% (concentrated on the 80% of residential customers who have lower bills), and planned to remove retail price caps in 2001, subject to the development of competition. Access prices for other suppliers wanting to use BT’s networks would continue to be regulated. As 1997 was the election year which brought the change from Conservative administrations to a New Labour government, the overall position at the time was well set out by the third report of the Trade and Industry Committee on Telecommunications Regulation published in March 1997 (HC254).

The Committee considered that competition had been actively promoted between 1984 and 1997, evidenced by the fact that over 200 licences had been granted to over 50 major telecommunications operators, the duopoly in the international facilities market had been ended in 1996, and the European telecommunications markets would be opening in January 1998, but the Committee noted that BT still continued to dominate the markets, as the following table showed:

Market shares of international and national business and residential calls				
	International calls		National long distance calls	
	Business	Residential	Business	Residential
BT	53%	81%	73%	90%
Mercury	23%	9%	19%	6%
Cable	2%	8%	2%	4%
Other	22%	1%	6%	0%

However, the level of investment in network infrastructure was so high that it suggested that these figures might simply represent the long time it takes to build the network and develop a customer base. Another important development was ‘number portability’, the ability of customers to keep their telephone number when they change operators, following an MMC

referral in 1995, which meant that BT was now required to provide portability to competitors on a reciprocal basis and at a charge determined by the Director General.

The Committee also noted the restrictions on public telephone operators (PTOs) concerning broadcast entertainment services, which had been imposed to encourage the investment needed for local companies to build their networks and become established. Given the investments now made the Committee recommended that such restrictions were lifted by 2002. The impact of technical change also had to be taken into account. Telephones to domestic properties were connected by copper wires, although the trunk system was now fibre-optic cable. It had been expected that a broadband network throughout the UK would be based on fibre. However, technologies such as Asynchronous Digital Subscriber Line meant that BT was now able to provide many interactive broadband services to residential customers over its existing copper network (the 'local loops'). BT's position was that the restrictions potentially distorted competition because it was unable to develop effectively and efficiently its network using digital technology if it is prevented from employing broadcast technologies.

One development which was to become a major policy for the incoming New Labour government concerned the role of the independent 'director general'. For both BT and British Gas, it can be seen why they might think too much regulatory power had been placed in the hands of one individual, given the pressure on them to be more competitive. In paragraph 30 the Committee said "BT has no objections to Oftel becoming more like a competition authority and having increased powers provided it was done by statute and there was a right of appeal to the courts. BT also argued for a system in which decision-making was the responsibility of a small panel rather than of one individual, similar to the American collegiate system, since this would de-personalise decision-making. We recommend that this should be a matter for consideration."

On price controls, given the decline in the price controls on the residential market, the main concern for the competing companies was the proposed new network charge cap, a control on the charge by one operator for carrying the traffic of another, due to come into effect from August 1997. The committee said "Oftel has proposed radical changes to the basis on which interconnection charges are set, from historic to long run incremental costs, and to the process of setting charges: in future BT not Oftel will set connection charges. The subject is currently under consultation and some very complex issues which are likely to have a profound affect on the structure of the industry will need to be resolved. We believe that our successor Trade and Industry Committee should examine this issue again in the next Parliament." (34).

The announcement in November 1996 that BT was proposing to merge with MCI, an American company, to form a new company called Concert plc was a major concern resulting in two recommendations if the merger went ahead. First, "We recommend that the government should make a commitment to safeguarding the interests of UK customers and should give appropriate policy direction to Oftel to have regard to the level of future investment in the domestic market." and, secondly, "We recommend that the Director General examine the potential consequences, on the development of competition, of a closer relationship between BT and BSkyB and take appropriate action, including, if necessary, requiring Concert to dispose of its interest in News Corporation (the parent company of BSkyB) or amending BT's licence conditions." (41 and 42).

Finally, the Committee reflected on the increasing convergence between the telecommunications and broadcasting industries, and recommended that the government should review the structure of telecoms and broadcasting regulation. It noted that the Consumers' Association had argued

for a single regulatory body for the communications industry, and that the Hansard Society had recommended enhancing the role of Oftel by creating an Office of Communications which would regulate infrastructure in telecoms and broadcasting, but leave the Independent Television Commission (ITC) to regulate issues relating to content. Given the format of the new independent economic regulators had been ‘Office of...’, the Committee had set out the possibility for a new Ofcom!

- British Gas privatisation and regulation

Following the privatisation of BT, the British Gas Corporation followed in 1986. In order to maximise receipts from privatisation and widen share ownership, the government embarked on a very effective marketing campaign, referred to as ‘tell Sid’, but in various ways the approach to privatisation and regulation differed from that of BT. Two sources of information are important: the First Report for the Energy Committee of the House of Commons (session 1985-86) entitled Regulation of the Gas Industry, and the annual reports of the new regulator, the Office of Gas Supply, the director general being James McKinnon. It is interesting to note that the Committee’s report made the following observation “Labour members of the Committee have made plain to the Committee throughout the enquiry their opposition to the privatisation of the gas industry, but they accept the need for this report.” (part VIII, 98).

The committee usefully set out the two types of regulation to be considered, first the regulation of standards, covering measurement, health and safety (issues which did not raise questions specific to ownership of an industry) and, secondly, economic regulation “designed to compensate for the absence of competition.” (part III). As British Gas was sold, like BT, as a single legal entity, the committee was concerned to consider the “four differing business activities – activities often separately owned in other countries; viz:

- * Hydrocarbon exploration and production
- * Gas transmission and distribution
- * Gas retailing
- * Appliance sales and after-sales service.”

Given BG’s effective monopoly in the first three of these businesses, and the possibility of cross-subsidisation across all four businesses, the committee was concerned to examine regulation of the gas market in three parts:

- 1) Regulation of prices to consumers
- 2) Regulation of competition in supply to consumers
- 3) Regulation of the offshore market (part IV).

A distinction was made between the ‘tariff market’ for smaller domestic customers taking less than 25,000 therms per annum, and the contract market which applied to large, industrial consumers, which would continue as before. The price control for the tariff market had to be formulated differently than that for BT because a large part of BG’s cost basis is the purchase of gas. This meant the formula had two variables in addition to the RPI: an efficiency factor X covering the non-gas costs per annum, and a Y factor covering the increase in the average cost of gas taken during the relevant year. This meant that real terms falls in gas supply prices would only occur if X was greater than Y. An additional element compared to the BT formula was that BG had to forecast the RPI increase for each year. This meant that there was a correction factor K each year and “this reduces or increases the maximum average price per therm in any subsequent year to recoup to BG or its consumer any over or undercharging in the previous year resulting from inaccurate forecasting.”

Another major difference concerned consumer representation because, although it was the role of the economic regulator to protect consumers, there already existed a National Gas Consumers Council (NGCC). The government therefore established a Gas Users Council (GUC) to replace it at privatisation, to be called the Gas Consumers Council, and the GCC and Ofgas played separate but complementary roles. The GCC had regional offices and the relationship developed with Ofgas was set out in the First Report of the Director General of Gas Supply (February 1987), stating that “Throughout these discussions it was accepted that gas consumers should refer their concerns to BG in the first instance. Where a consumer remains dissatisfied, the regional office of the GCC would normally be the next port of call. The GCC has a duty to notify Ofgas as soon as it is apparent that the Director General’s enforcement powers may be exercised in any particular case.”.

As both BT and BG were sold intact, then both Oftele and Ofgas faced similar regulatory problems; how to develop effective competition, and how to deal with issues of social equity. Following privatisation there was an increase in the number of disconnections from gas supply, and in its annual report 1987 Ofgas stated that “When statistics on the number of disconnections became public knowledge we received representations in every region that the increase was directly attributable to a change in BG policy operated subsequent to privatisation. This is not a view to which I subscribe. I suggest that increased consumer inability to pay for gas has been the cause.....It seems that the largest number of disconnections arise in those parts of the country in which poverty and deprivation are most prevalent.” (page 5). The problem of ‘fuel poverty’ continued in public debate for many years and Ofgas applied significant pressure on BG to be flexible with customers and find acceptable ways to pay for gas consumed, including the right to have a prepayment meter.

Ofgas proved to be very firm in its intention to see effective competition develop and the opportunity came very quickly because in 1988 BG had been referred to the Monopolies and Mergers Commission (MMC) by the Office of Fair Trading (OFT) on matters affecting the contract market for gas supply. In its annual report for 1988 Ofgas reported on the MMC report published in October 1988, stating that although the MMC “identified and described its finding that British Gas had systematically discriminated against certain types of customer, its main thrust was to create conditions in the market which would enable competition to develop among gas suppliers.....it identified the need for transparency in British Gas industrial pricing policies, the pre-requisite that gas should become available for competitors and that they should be fairly accommodated on the British Gas transmission and distribution systems.” The MMC report was concerned whether Ofgas had sufficient powers to enable competition to develop, and although Ofgas stated that sufficient powers did exist, “the credibility of such a statement was adversely affected by the fact that no practical example could be described.” From this point on Ofgas sought to give directions to BG where a competitor was seeking a ‘common carriage’ contract. Throughout 1987 and 1988 Ofgas also negotiated with BG the preparation of disaggregated financial data to see the financial results of the different segments of their gas supply business. In 1989 the MMC ruling meant that the government brought forward plans for other suppliers than BG to buy gas from north sea producers. This meant that BG only had ‘first refusal’ on buying up to 90% of new gas rather than 100%. The 90/10 rule facilitated the development of competition to BG.

The erosion of BG’s monopoly power continued in 1990 when Ofgas ruled that a rival company could supply its own customers through the national gas grid and allowed two companies, Agas and Quadrant Gas, to enter the market as suppliers with direct access to gas reserves. Progress was still slow, however, and Ofgas warned BG that unless competitors can command some 30%

of the gas market by 1993, it might be necessary to consider restructuring the gas industry on the lines of the electricity industry, with a separate grid company. BG must have recognised that things had changed from being a nationalised industry, and began to diversify into exploration and production, as well as expanding internationally. The tone of the relationship between the Director General and BG was well reflected in the DG's statement in the 1989 annual report:

"It is over three years since British Gas was privatised and during the great part of that time we have found the attitude displayed by the company towards change to have been somewhat negative. This has been a matter of concern to us. Time has been lost and goodwill dissipated as a result of the general approach adopted by the management of the company. Goodwill can be regained but it is never possible to recover lost time. In short, the initial opportunity for British Gas to make an impressive start in its new role has been lost.....Thus the year 1990 will be of critical importance as it is likely to demonstrate whether British Gas has really changed its approach.....It would be a further positive signal if British Gas were to take active steps to accelerate the process of competition and to banish the gripes which have long been expressed about its status as a monopolist."

The pro-competition approach continued in 1990, shown by Ofgas's annual report in which the DG's statement had a long section on the acceleration of competition in the industrial market, but what is interesting is their need to set out on page 1 the position of Ofgas in relation to BG, stating "It is interesting that one basic factor affecting Ofgas has not changed since 1986. As far as our main activity was concerned, we were then faced with having to deal with a total monopolist and that is the position today. On the other hand the external perception of our powers has changed to a dramatic extent. In 1986 Ofgas was written off as being a 'toothless watchdog', whereas today we are beginning to be portrayed as savage, dangerous and out of control. Neither of these two descriptions is related to the reality of the situation.....We believe the main motivation of Ofgas has been the desire to see the gas supply industry operate on the basis of a normal market place and that has been applied equally to the tariff segment and the industrial and commercial sector."

In 1991 the price cap formula for tariff customers was revised, in part to take a more 'green' approach, and secondly to expect BG not to be given complete pass through of gas purchase costs in the Y factor, but to recognise that BG has some market power in its purchase contracts, and that BG should be given an incentive to use it. The costs which could be passed through were energy efficient schemes approved by Ofgas, thereby encouraging BG to consider energy efficiency options as well as energy supply. The new formula to replace $RPI - 2 + Y$ was:

$$(RPI - 5) + (\text{Gas Price Index} - 1) + (\text{Allowable energy efficiency expenditure } E)$$

The other part of the story, which had been an issue for BT and BG ever since they were privatised intact, was ongoing pressure for structural reform. This built on the agreement reached in 1991 with the Office of Fair Trading that gas should be released to competitors in the contract market to achieve a 40% market share, and to arrange for a clearer separation of gas transportation and storage from its other businesses. As the regulators continued to try and develop competition, the issue of fair access to the networks grew in importance, which brought into question the sustainability of integrated ownership. By 1992 British Gas had again been referred to the MMC. British Gas was concerned that the allowed rate of return had not been correctly calculated and that the 1991 formula for the tariff market was being affected by the changing position in the contract market, given the government's reduction in the threshold on British Gas's statutory monopoly for supply to customers from 25,000 to 2,500 therms per year.

British Gas had concluded that a fair rate of return for its transportation assets should be around 6.7% whilst Ofgas suggested that it lies in the range of 2.5 – 5.0%. The dispute over these led to British Gas asking the President of the Board of Trade in July 1992 to refer the gas tariff market to the MMC, saying in a letter that “Sir James Mckinnon (Director General of Ofgas) is not prepared to face the inevitable prospect of increasing transportation prices, but would prefer to leave that uncomfortable task to the MMC.” Ofgas simultaneously referred the transportation and storage market to the MMC, and made a parallel reference on the gas tariff business in August 1992 in order to allow the MMC more flexibility in framing its recommendations.

Ofgas had decided that British Gas must be broken up to facilitate competition, and published its own proposals in December 1992. In the annual report for 1993 the DG said “Let there be no shadow of doubt that our reference was made on the strategic issue of ownership.” (p7). Accounting separation had been developed but Ofgas recognised that the key to BG’s monopoly was its ownership and control of the gas transportation and storage system, concluding that “for it simply to be separate, and to allow it to be owned and controlled by British Gas, cannot serve the objective of achieving real competition.” Three reasons were:

- * Conflict of interest between promoting third party access to the transportation business and giving preference to its own gas supply business;
- * The scope for discriminatory pricing arising from the allocations of joint and common costs;
- * No guarantees that commercially sensitive information given to the transportation unit by competitors would not be used by the supply side of British Gas to protect its dominant position.

The MMC reported its findings in August 1993 (Gas Cm2314 and British Gas plc Cm2315). With four references having been made to the MMC (two from Ofgas under the Gas Act 1986 and two from Department of Industry under the Fair Trading Act 1973), the MMC in practice treated them as one inquiry. The four main conclusions were:

- * British Gas’s operation of its pipeline and storage systems may be expected to operate against the public interest;
- * The transportation and storage businesses should be regulated to stop discriminatory pricing and British Gas should ‘divest’ itself of its gas trading business;
- * In recognition of the government’s reduction in the tariff market threshold, the price cap formula in the tariff market should be relaxed from RPI – 5 to RPI – 4;
- * The tariff threshold should be further reduced to 1,500 therms in 1997, and the monopoly should be completely removed some three to five years after divestment of British gas’s trading activities.

In December 1993 the government announced that it would not be accepting the MMC’s recommendation to require the break-up of British Gas. This was partly due to concerns from consumers about higher prices for domestic customers as cross-subsidy was withdrawn due to more competition in the contract market; investor resistance to the right of government to alter the structure of British Gas (given it was privatised with a 25 year authorisation licence); and political concern about the international competitiveness of one of Britain’s largest companies. The government also did not accept Ofgas’s suggestion that the British Gas supply business be broken into 12 competing regional supply businesses.

In 1993 a new Director General came to Ofgas, Clare Spottiswoode, and it was hoped that the previously acrimonious relations between BG and Ofgas would be mellowed. They worked together in 1994 to achieve real separation between the transportation and storage businesses (called Transco) and the gas supply (public and contract) businesses. BG was required to have

different staff in different locations for the two businesses, with firm regulatory rules on the transfer of information, and Ofgas appointed a compliance officer. Ofgas announced its proposals for price controls on gas transportation and storage in August 1994. This reflected the fact that there was a potential conflict between price structures which promote competition and those which reflect costs, and so 'facilitating effective competition' was therefore made subordinate to the objective of 'cost reflectivity'.

The physical separation of the two businesses had not, however, taken away the widespread concern that a vertically integrated monopolist was not the best way to proceed, particularly given the developments in privatising the electricity and rail industries from 1990. In February 1996, ten years after the privatisation of British Gas intact, British gas announced that it would demerge its company into separate listed companies – Transco International and British Gas Energy. The decision to demerge was taken in the interests of efficiency, regulatory stability and business focus, and in February 1997 the formal demerger took place with the new corporate titles: Centrica for the domestic trading business and BG plc for the rest, including Transco International and the global gas business of production and exploration.

- BAA privatisation and regulation

The privatisation of the British Airports Authority (BAA) in 1997 raised some similar issues to those faced when privatising BT and BG, but it was also significantly different. For a start BAA did not own all of the UK's airports, airports have significant environmental impact as they tend to be near large centres of population, and airport policy has to take account of the fact that international air travel is already affected by international treaties, as well as the dramatic increase in air travel since the second world war, both leisure and commercial. The position, and protection of, the privatised British Airways (BA) had also to be taken into account, and any development of airport competition would have to take account of the complexities of managing landing and take-off rights (slots) in terms of runway capacity, as well as terminal capacity to receive arriving and departing passengers. Such complexities meant that airports policies of successive governments before 1979 had given rise to almost unceasing pressures, criticism and public controversy, particularly the proposal to build a new London airport on the Thames estuary at Maplin, which was abandoned in 1974. From 1981 to 1983 public inquiries had been held into proposals to develop Stansted and Heathrow airports because of the need to expand capacity. The government's response was to publish a white paper on Airports Policy in 1985 (cmd 9542) which sets the framework for the structure, ownership and regulation of UK airports.

The pattern of ownership of airports reflected that planned in the 1961 white paper on Civil Aerodromes and Air Navigational Services (cmd 1547). The British Airports Authority, a statutory corporation, was established by the Airports Act 1965, and owned three of the London area airports – Heathrow, Gatwick and Stansted – and the four major Scottish airports – Edinburgh, Glasgow, Prestwick and Aberdeen. These seven airports accounted for nearly 80% of the international passengers handled by all the UK airports. Most of the other major airports were owned by local authorities, with a few privately owned airports, and the Civil Aviation Authority (CAA) owned and operated for social reasons a number of small aerodromes in the Highlands and Islands of Scotland serving remote communities. The main airport in Northern Ireland, Belfast International, was wholly owned by the Department of the Environment, Northern Ireland. The white paper stated that "the government is committed to converting as many as possible of Britain's airports into private sector companies as part of its policy of reducing the role of the state." (9.5).

Expecting airport privatisation to increase competition between airports was not realised because the government decided to privatise the BAA as a whole (as with BT and BG), stating that “although selling off its seven airports separately would have reduced BAA’s dominance, this option was found not to bring substantial real advantages, and to have some important disadvantages.” (9.9). This was because the volume of business switching between airports to enjoy lower competitive landing charges would be very small, given that airlines operating scheduled services aim to maximise their carrying of connecting passengers, so that the revenue effect is far more important than the cost of landing charges. The evidence for this was that Heathrow had maintained its dominance as a traffic hub even though it had higher landing charges. However, in order to inhibit cross-subsidy between BAA airports (such as Stansted being subsidised from the profits of other BAA airports), the government decided that “Stansted and the other BAA airports should be formed as separate companies, under one holding company, with each of the seven airports producing separate accounts.” The government thought that this would ensure transparency, effective management control, and a pricing system that would not distort the demand for air transport between the airports in the London system and airports in other parts of the country.” (9.8). Privatising BAA intact also meant that the government would retain maximum flexibility in the administration of government aviation policies, covering route licensing, traffic distribution and airport development, while at the same time allowing BAA plc to “expand into non-airport activities from which the BAA as a nationalised industry had been excluded.” (9.6). The reason for not selling BAA’s London airports separately from its Scottish group was that it would do nothing to increase competition, and also lose significant economies of scale of operations between Scottish and London airports, including interchange of staff.

For local authority airports the government felt that they would be better run as Companies Act companies with the shares held, initially at least, by their local authority owners. The government felt that “this would bring several advantages: a formal management structure which would define directors’ responsibilities more clearly...and an ownership structure which is flexible and can facilitate changes of ownership within the public or private sector.” (9.15). The government therefore proposed to introduce legislation which would give the government power to require local authority airports with an annual turnover of more than £1m to be formed into companies. The government would “encourage local authority owners to involve the private sector as concessionaires and contractors in the development and running of their airports...and give the local authorities powers to dispose of some or all of their shares in the new airport companies to the private sector as well as to other local authorities....something that would allow new capital for the financing of airport development which would be required to meet the growing volume of traffic.” (9.16).

In terms of regulation, privatisation meant certain new measures had to be put in place in addition to the regulation of aviation already in force. The regulation of security (anti-hijacking and anti-terrorist measures) already applied to all airports regardless of ownership since the Aviation Security Act 1982, but this had to be taken into account if there was to be price control applied to the privatised airports. The formulae would require a new +S to allow the pass through of any additional costs which the regulation of security imposed within the price control period. To give the airports the incentive to manage such new regulations cost-effectively, only 95% of the cost was to be allowed to be passed on to customers. The Chicago Convention sets out the UK’s international obligations which require that airports open to public use by UK registered aircraft are available to the aircraft of other states without discrimination. This had a particular impact on airports and regulators that might wish to use opportunity cost as a means of pricing landing charges, as “the UK also has an international obligation to use its best efforts

to ensure that airport charges are related to costs (accounted) and allow only reasonable profits.” (10.6).

The regulatory model adopted for the privatisation of BT and BG, involving a new independent Director General, was, however, not followed for the privatisation of BAA. The Civil Aviation Authority’s existing responsibilities for the UK aviation industry were extended to cover the regulation of airport’s traffic charges. The government adopted the price control system in place for BT and BG but not the process by which the price control proposed was offered to the regulated company which could then decide to accept or appeal to the MMC. The system for airports regulation would be that “towards the end of each five year period the Monopolies and Mergers Commission would review the airport company’s affairs, including its cost structure and efficiency, and would report to the CAA; appropriate steps would then be taken to remedy practices found to be adverse to the public interest and, on advice from the MMC, the CAA would lay down an appropriate formula to govern traffic charges for the next period.” (10.10). So, unlike the first two privatisations with one economic regulator and an appeal mechanism, we now had two economic regulators running in parallel, and with no route for appeal.

The price control formula administered by the CAA was a revenue cap in that it specified that the maximum annual average revenue per passenger at Stansted, Gatwick and Heathrow together, and at Gatwick and Heathrow individually, could not increase in any year from 1987 to 1992 by more than the rate of inflation less 1%. The first review carried out by the MMC reported in June 1991 and recommended that the price control should be tightened to RPI – 4% for the next five years. This should be abated to RPI – 1% for the last two years if Terminal 5 at Heathrow was to start in 1995/96. The cost pass through of additional security costs was recommended to increase from 75% to 85%, to be recovered one rather than two years in arrears. The CAA’s reaction to this was to propose a much tighter price cap of RPI – 8% for the full five years, primarily because it felt that the cost of capital for BAA was nearer 7% rather than the 8% recommended by the MMC. The CAA however changed its position under the weight of argument from BAA and the lack of support from BAA’s customers, the airlines. The necessary rate of return was increased to 7.5% and the five year formula was profiled to take account of the impact of a major capital project such as terminal 5. The profile adopted for 1992 to 1997 was RPI – 8% for two years, RPI – 4% in the third year, and RPI – 1% for the last two years. This approach eliminated the suggestion that – 8% was a sustainable annual productivity gain in the long run. The security cost pass through was increased to 95%.

By 1994 international and domestic competition issues had come to the fore for BAA. Luton airport applied to the EC for a judgement on whether BAA’s charges on Stansted were too low, and consequently an abuse of their dominant position under articles 85 and 86 of the treaty of Rome. The preliminary findings of arbitration following complaints by Pan Am and TWA about landing charges at Heathrow suggested that perhaps there should be amendments to the international agreements which would make the test criteria more consistent with the incentive price cap system used in the UK, rather than being based on strict annual rate of return limits. The Department of Transport had also started a review of the regulatory framework for airports, which might result in the CAA, like the other utility regulators, taking the lead in periodic reviews, rather than the MMC. In 1995 the Government announced that it intended to put the CAA on the same footing as other independent economic regulators. In preparing for the 1997 price review the CAA had to face the difficulty of how to set the controls when the impact of terminal 5 had to be taken into account, and with the ‘single till’ approach in place, which incorporates income from commercial activities at airports along with landing charges levied on airlines. The CAA announced that it would work on the assumption that terminal 5 would be

built, and have an ‘interim’ review of the price cap if it wasn’t. This operated the opposite way round from the other economic regulators, who relied on interim review cost pass through procedures once the new expenditure was actually in place (one example being Ofwat in relation to new expenditure caused by European directives). The debate about competition in aviation focused to a large extent on the ‘the economics of airport slots’. Airlines want to compete but require the ability to take off and land at airports which are suitable to them. Whilst airports do not compete significantly in terms of landing charges, the allocation of landing ‘slots’ is very significant to the airlines ability to compete.

To get a clear view of the developments and problems facing the regulation of airports since 1986, it is best to move forward to 2006, when the Transport Committee of the House of Commons published its 13th report of session 2005-06 on *The Work of the Civil Aviation Authority* (HC809). In nine sections it examined: Management of the CAA; European Aviation Safety Agency (EASA); Performance of the CAA; The CAA’s resources; Economic Regulation of Airports; Airspace Regulation and Aviation Sustainability; and General Aviation. The section on economic regulation of airports is important because it confirmed the weaknesses which resulted from the different institutional framework it had to operate under compared with the other utility regulators. First, whilst the Airports Act 1986 made the CAA responsible for the economic regulation of any airport with an annual turnover of more than £1m, in practice economic regulation focused only on those airports designated by the Secretary of State, and these were four: The three London airports of Heathrow, Gatwick and Stansted, and the local authority owned Manchester airport. Designation meant that the CAA had to set a price cap, but before it could do that it had to make a reference to the Competition Commission, asking it to judge the appropriate maximum limit on airport charges for the next five years, and to report on whether the airport had pursued a course of conduct contrary to the public interest in the last five years. The CAA had to have regard to the CC’s findings but was not bound by them.

The CAA and the CC had an important exchange on the use of a ‘single till’ or a ‘dual till’ approach to setting a price control. In 2003 the CAA had proposed a dual till approach, which meant airport charges were calculated with reference to the costs and revenues associated with providing aeronautical services alone, and excluding the commercial revenues from retail shops at airports. This would eliminate the subsidy on airport charges resulting from including retail revenues, and help to price airport charges more in line with opportunity cost. However the CC considered that this would result in too high airport charges, and be inconsistent with the UK’s international obligations, so concluded that the single till was most appropriate. The CAA accepted the CC’s recommendations.

The airlines could be critical of the CAAs approach to price reviews, sometimes arguing that the CAA was too narrowly focused on economic theories, rather than the pragmatic reality of running an airport. Their response was that “the designation of airports by the government rather than by the regulator itself, the mandatory setting of price caps for five years, and the automatic referral to the Competition Commission, were all factors that limited the CAA’s flexibility to tailor its regulation to the evolving airport’s market.” (122.). The Committee heard much evidence that it would be sensible to de-designate Manchester and Stansted airports because they were not in a position to abuse their monopoly power, and therefore the costs imposed by a price review were unnecessary, and if necessary any abuse could be dealt with by the CC. The automatic referral to the CC came in for substantial criticism because it seemed like a ‘duplicated’ review process, given also that the CAA was the overall regulator for the aviation industry, and therefore should be expected to have more expertise than the Competition Commission. The Chairman of the Better Regulation Commission noted that the dual process

eliminated a route of appeal for the regulated, and considered it to be an “odd situation”, an anomaly which the government should look to correct. (129).

On the positive side, by 2005 the CAA had introduced the concept of ‘Constructive Engagement’, whereby airports and airline customers would be given the opportunity to reach agreement on a number of issues, with the CAA acting as facilitator rather than the decision maker. In this way “the normal business of commercial airport/airline interaction should be reinforced by the regulatory process, rather than interrupted by it.” (134). The issues that were to be tackled by this method were:

- * Volume and capacity requirements
- * The nature and level of service outputs
- * Opportunities for operating cost efficiencies
- * The nature and scale of the investment programme
- * The efficient level of capital expenditure associated with that programme
- * The revenues from non-regulated charges to airlines
- * The elements of service quality and investment to which specific financial incentives should be attached.

The nature of these makes it clear that engagement with those who were closely involved with the business would give the regulator an opportunity to reduce the burden of the price review process, and where difficulties did arise, particularly in relation to Stansted, it was a case of ensuring that all parties were incentivised to participate fully. (136-139).

How best to deal with the capacity of London airports was a policy issue which arose throughout the period 1984-1997. The white paper on airports policy in 1985 in considering ‘traffic distribution policy’ stated that “the government proposes to introduce legislation whereby the CAA would, at the secretary of state’s request, have the duty of developing and promulgating a scheme for allocating scarce airport capacity.” (11.17). One problem was that competition between airlines, which was increasing, was restricted by the traditional method of allocating scarce landing and take-off slots. This was left to the ‘Airports Scheduling Committees’, made up of the airlines using an airport, and operated on a give and take basis within guidelines agreed internationally between airlines. In practice the basic working rule was to respect ‘grandfather rights’, that is, those airlines already operating had first claim on the slots that they were already using.

To facilitate the development of competition the government looked to the CAA to develop an alternative system, but recognised that trying to achieve this by a market-clearing system of airlines ‘bidding’ for slots, and disregarding grandfather rights, might be a difficult policy issue. It summarised it by stating “such is the potential demand at Heathrow and Gatwick that such a system could not be expected to work without raising the general price of slots at those airports well above the airport’s long run marginal costs. This would not be compatible with our international obligation to ensure that airport charges are cost related. Moreover, its likely effect would be to reinforce rather than weaken the hold of the present users of slots. Notwithstanding the value to Heathrow and its users of its unrivalled opportunities for ‘interlining’, there is a danger that the domestic services which form part of that value would not be able to withstand the competition from the international services, whose larger aircraft would on the whole be better able to afford the consequent higher landing charges.” (11.7).

The complexity of allocating slots in an economically efficient way were well set out in a 1993 article in the Topics series of the National Economic Research Associates (NERA) entitled The

Economics of Airport Slots by Ian Jones and Ivan Viehoff. 1993 had been an important year because the transport ministers of the European Community had negotiated a major air service liberalisation package, which made the need for effective competition even more important. This was because, as the authors concluded, “the market failure in the market for airport services is a worldwide phenomenon, deserving the urgent attention of the relevant authorities in Europe, North America and the Far East.”. Their recommendation was to improve the functioning of the market for airport services by a combination of policies which both allow the establishment of ‘property rights’ to airport slots and reduce the present cross-subsidy from other commercial activities to airport charges. Unsurprisingly, given growing demand for air travel, one way to avoid the problems of scarce airport slots was to focus on the need for new runways and terminals at the major UK airports.

- Water Authorities in England and Wales, Privatisation and Regulation

The privatisation of water authorities in 1989 provided some distinctive contrasts with the three utilities that had gone before: BT, BG and BAA. These are well set out in two reports published in 1986, which started the process towards privatisation in 1989. The first was the government white paper which followed the decision to privatise the 10 water authorities in England and Wales: Privatisation of the Water Authorities in England and Wales, February 1986, Cmnd 9734, which was supported by the second: Economic Regulation of Privatised Water Authorities, Report to the Department of the Environment, SC Littlechild, January 1986.

In contrast to Professor Littlechild’s first report on BT, which had focused in the longer run on competition developing and therefore reducing the need for price controls, this report focused on the fact that “First, the water industry is largely a natural monopoly, and second, privatisation will involve not one but ten water authorities.” (1.1). It followed therefore that “Competition is not an alternative to regulation of the water authorities, but it is an important complement which can facilitate the regulator’s task and provide added protection to customers.” (1.6). In particular, competition in the capital market was seen as an appropriate pressure on a water company to run its business efficiently and innovatively, for fear of a takeover bid, but this would be undermined if the government chose to prevent takeover with a ‘golden’ share. Comparative information between the ten water authorities would be useful in setting price controls but “This, and the burden of negotiating with ten authorities, means that regulation needs to embody as much uniformity as the different circumstances of the authorities permit.” (1.4). Such different conditions might suggest the need for up to ten different values of X in applying the RPI – X tariff formula, but the recommendation was that there “is a stronger case for a uniform X across all authorities. This would avoid accusations of discrimination, facilitate operation of the ‘industry yardstick’ for revising X, and minimise the burden of licence negotiations.” (1.18). The focus on levels of customer service and environmental standards was particularly important for a water industry concerned with both water supply and sewage disposal. It was made clear that “Regulation must encompass not only prices but also levels of service. The two aspects must be considered simultaneously.” (1.2). This would also mean there should be “A licence requirement periodically to publish long term plans to meet future demand which would provide useful reassurance and information concerning water resources.” (1.20).

The government’s policy objectives for privatising the water authorities were the same as previous privatisations (converting the water authorities into water service public limited companies – WSPLCs), but account was specifically taken in the white paper of the organisational structure of the water industry and the need to manage the environment. For the

latter “The environment will benefit from a new system of government approval for the objectives set for each river and estuary, defining the standards to which each is to be maintained and improved. Fair and effective procedures will be maintained to regulate all discharges of effluent or other matter to rivers, estuaries and coastal waters, and the arrangements recently introduced to ensure public access to information about their quality will be further enhanced.” (8.). The water authorities would be privatised on the basis of their existing boundaries because “The principle of integrated river-basin management – a single body controlling water and sewerage in each river catchment – has worked well since it was introduced by the water act 1973.” (11.). By this the government felt that “The proposals in this white paper represent a significant extension of the government’s successful programme of privatisation and will put a major national industry onto a new and dynamic basis, while protecting the interests of all who use water and safeguarding public health and the environment.” (15.).

The nine water authorities in England were South West, Wessex, Southern, Thames, Anglian Severn-Trent, Yorkshire, North West and Northumbrian. Welsh Water covered the majority of Wales. Over 99% of the population of England and Wales is connected to a public water supply and 96% to a public sewer. Before the 1973 act, there had been nearly 1,600 local undertakings responsible for water services, so integrated river basin management greatly simplified the overall structure. However twenty eight statutory ‘water supply only’ companies remained and supplied about a quarter of the water to the population of England and Wales, and some water authorities discharge their water supply functions in part of their areas through the statutory water companies. The water authorities, apart from supply water and sewerage services, also regulate other users of water through a licensing system for the abstraction of water from rivers and other sources. It is interesting to note that in 1984 public water supply was 3,100 million gallons per day, but in addition 6,600 million gallons per day were abstracted for non-domestic purposes, with about a third of this used by the Central Electricity Generating Board for cooling at power stations. They are also responsible for environmental conservation, fisheries, navigation, flood defence and land drainage. Given the local authority involvement prior to 1973, this was maintained by the large boards which were established for water authorities, with a majority of local authority representatives. However, this was changed by the Water Act 1983, which provided for smaller, more business-like, boards and where all the members were appointed by government ministers. This was taken further in the white paper when it said “The Secretary of State will select new members with an eye to the future as appointments expire. He will no longer consider himself committed to appoint members nominated by local authorities or who possess local authority experience.” (102.).

The statutory water company model was considered as an alternative to privatising the water authorities as PLCs. The financial controls on statutory companies are written into the companies’ statutes, which limits the maximum amount of dividends that can be paid, the size of reserves, and the amount of balances that can be carried forward from one year to another. In essence it is closer to the rate of return regulatory model than price control. The government concluded that “To establish statutory companies is an out-dated method of forming a business and powers necessary to protect investors are already provided by general company legislation. If necessary, additional safeguards can be made, as in previous privatisations, through government ownership of a ‘special share’ for purposes defined in the WSPLCs’ articles of association.” (42.). The government’s aim was that statutory water companies could convert to PLC status and fall under the same financial regulation as would apply to the WSPLCs. Another model the government considered as an alternative to full scale privatisation was a system widely practised in France whereby the water authorities would be required to ‘franchise

out' the whole of their operational functions, for which potential franchises could bid competitively. This model was rejected in part because "the division of responsibility between owner and operator would be prejudicial to a properly planned programme of maintenance, renewal and replacement." (49.).

The government decided, as in previous privatisations, that regulation should be the responsibility of a Director General of Water Services, stating that "the Director General's principal duty will be to safeguard the interests of the consumers and ensure at the same time that the companies have every incentive to perform well in the interests of their shareholders." (57.). The costs of the director general and his staff (Ofwat) would be met from the fees charged to the water companies for granting a licence. One important development was that the DG would also be responsible for setting up and maintaining new consumer consultative arrangements, again to be paid for from the licence fee. Formal arrangements to represent customers' interests had been put in place for the first time by the 1983 act. However, privatisation gave the opportunity to strengthen the responsibility of consumer bodies by making them regional consultative committees. The white paper said that "The regional bodies will be independent of the water authority they monitor and will report to the Director General. In addition to their existing duties, the government wants them to investigate specific complaints, including complaints of maladministration, on behalf of consumers, in cases where they have not received a satisfactory response from the WSPLC themselves. They will have an ultimate right of complaint to the Director General. Accordingly the government intends to remove water authorities from the jurisdiction of the Commission for Local Administration (the local ombudsman). The DG's office will, however, be subject to investigation by the Parliamentary Commissioner for Administration (the parliamentary ombudsman)." (77.).

Having accepted the argument in favour of setting uniform standards and price controls throughout the industry, rather than controls tailored to each WSPLC, the government still faced the fact that section 30 of the water act 1973 stated that charges should be cost-related and non-discriminatory between classes of customers. Most domestic consumers were charged for water services on the basis of local authority rateable values (a measure of comparative property values). This meant the government would need to discontinue the use of rateable value as the basis for water charges. All industrial, and many commercial customers, were already metered for consumption, but to extend this to domestic customers raised the issue of the cost of installing water meters, and who would bear the cost. Also, in cost-benefit terms, it would depend on whether domestic water metering would lead to a significant change in the volume of water consumed. For this reason the government decided that the legislation should permit domestic water metering trials on a compulsory basis.

Given privatisation had occurred by 1989, the most important summary of that privatisation was the report by the National Audit Office in February 1992 entitled Department of the Environment: Sale of the Water Authorities in England and Wales, HMSO 256. The complexity of the sale was summarised as follows: "The simultaneous sale of ten separate companies on a single day was unprecedented in the United Kingdom. The restructuring of the water industry prior to the sale and the need to put in place regulatory arrangements which would apply not only to the privatised companies but to statutory water companies already in the private sector added significantly to the complexity of the task. The sale was further complicated by the need to ensure that the new companies would be able to finance a major programme of capital investment, the ultimate scale of which was far from certain and which would allow them to meet the existing water quality standards and the European water quality directives. The timetable was, moreover, necessarily constrained by the limited opportunities to

launch such a major flotation on the stock market, especially in view of the electricity flotations that were shortly to follow. Thus, the sale posed a wide range of difficult and unique problems for the Department and the results have to be judged against the scale of that task.” (4).

For regulation the Water Act 1989 had created three regulators: The Secretary of State for the Environment (to oversee drinking water quality), The National Rivers Authority (to manage pollution and environmental control) and the Director General of Water Services (for economic regulation including pricing). The new arrangements applied equally to the 29 statutory water companies. An important change from the earlier documents was that the K factor to be applied in the RPI + K formula was not to be a national average but an element calculated separately for each company to take account of their particular operational and capital expenditure needs. The DoE had “realised from an early stage that K would need to be positive given the scale of investment requirements in the early years.” (7). In order to have a balance between protecting customers from too high an increase in water prices but enable the privatised companies to finance their investment obligations, the government decided that initial Ks need not average more than 5% points, as long as the companies had capital structures suitable for their investment requirements. The RWAs had debts to the National Loans Fund (which cannot lend to private sector companies) of some £5 billion, and this had to be written off. The government established a capital structure for privatisation in which little or no debt was created and a cash injection of some £600 million was included. It was decided that there should be a 100% offer for sale with an individual share offer to the general public in the UK and a special share package to be made available to UK institutions and overseas investors. The water shares were priced at £2.40, which gave an aggregate capitalisation value for the 10 companies of £5,239 million.

The sale was judged to be a success, being 5.7 times oversubscribed, and this was in spite of the following constraining characteristics (29):

- * They were new trading companies and had no trading record;
- * They were subject to three kinds of regulatory supervision;
- * The RPI + K formula had been set for ten years;
- * The companies faced strongly negative cash flows over the medium term;
- * Substantial reliance would be placed on the economic regulator accepting ‘interim’ adjustments to K if additional burdens were imposed on the industry.

One of the first things that the new DG announced after privatisation was that he would take up the option provided in the 1989 Water act to carry out a review of the K factors after five years.

The year 1994 was an important one for two reasons. The normal five year cycle for reviewing the price controls had been chosen for the water industry, and the government’s ‘golden’ shares had been issued for all the ten water service companies for a period of up to five years (December 1994). The reason for the golden shares was that the water companies had to secure a number of environmental objectives following privatisation, and a period of stability was considered necessary. The golden share for Welsh water remained in place beyond 1994 because Welsh Water reflected a national identity for Wales, which would not be the case for the nine English water companies. A large amount of merger and takeover began from 1994 but within a clear policy framework for dealing with mergers. There was a mandatory reference to the Monopolies and Mergers Commission where the combined value of the assets in the merger exceeded £30m. The reason for the reference was that the Director General’s ability to compare water companies would be undermined by the loss of a comparator, and this would be against the public interest, unless the MMC could find that the merger resulted in a significant reduction in charges and that it was more able to move towards the frontier of efficiency. The first

takeover was that of Northumbrian Water by Lyonnaise, agreed in November 1995, where water price reductions of 15% spread over six years had been agreed (a deal negotiated by the DG).

One important development was the change from vertical integration in an industry to horizontal integration across utility industries. North West Water combined with the North West Electricity Board (NORWEB) in November 1995 to form what is now known as United Utilities. Licence amendments had to be agreed to ensure the regulators had access to the necessary information on each industry. In a similar way Welsh Water combined with the South Wales Electricity Company (SWALEC) in January 1996 to form what is known as Hyder. Scottish Power was able to agree a takeover of Southern Water by June 1996, having been successful over a bid by Southern Electric. Some of the water only companies also started to merge, as was the case with East Surrey and Sutton Water which was agreed in March 1996. The challenge of regulating privatised companies which might now be the subsidiary of a parent group meant licence amendments were made, including:

- * Appointed business to be conducted as if it were substantially the whole business and as if it were a separate public limited company;
- * Additional certification on compliance with 'ring fencing' arrangements;
- * No loans to parent or financial guarantees without the DG's consent;
- * The parent should refrain from actively causing the core business to breach its obligations and duties;
- * Ensure the board of the core business has two non-executive directors to strengthen protection of customer interests;
- * Core business's dividend policy should not impair its ability to finance proper carrying out of its functions.

The focus on improving competition in water services was less of an issue than it had been in the other utilities post privatisation, and this was reflected in the fact that the DGWS had been given a duty to 'facilitate' effective competition, not 'promote' it as in the other utility sectors. Proposals on common carriage focused mainly on the large industrial market, where there were some 600 companies with demand in excess of 250 megalitres a year. More political attention was paid to the issue of availability of water and water resources more generally, given the severe drought which affected much of the UK from the spring of 1995. In southern England and the central Pennines the rainfall was below 70% of the long term average. Public awareness was increased by the Environment Agency (previously the National Rivers authority) publishing a series of drought reports, and by the Water Services Association issuing regular reports to the press. This had some influence on reducing existing demand for water, but the longer term issue concerned the expected growth in demand for water. To increase the security of supply would require investment, and this would then be an important feature in forthcoming price cap reviews. It would not be sufficient to rely entirely on a variety of demand management measures such as subsidised water efficient appliances, selective water metering, leakage reduction and incentives on water companies to maintain supply capacity by introducing customer compensation penalties if water services are interrupted. The political pressure was most intense in relation to water leakage, where the Labour opposition pressed hard for mandatory leakage controls, notwithstanding that the optimal, or economic, level of leakage is hardly likely to be zero (the optimal level depends on the marginal balance between the costs of leakage reduction and the costs of providing extra water resources). The outcome was that in October 1995 the WaSCs committed themselves to reduce leakage over a ten year period to levels that are technically feasible and economically sensible by standards of best international practice.

Another political development which occurred in the years following privatisation was the response to the reported profits being earned. The basic model was that shareholders received the profits from outperformance for the period of the price review, and the regulator passed them on to customers at the next price review. The sustainability of this was however put under some pressure by the Labour party calling for a 'windfall' tax on utilities and by Ian Byatt, the water DG, suggesting that benefit sharing should not be too formulaic, and that if outperformance was delivered to customers early, then this could be given credit at the next price review. Some companies did start to share their performance with customers on an accelerated basis, such as South West Water's £10 rebate in May 1996. Another more technical problem which emerged was that the DG appeared to have increased uncertainty for the companies in relation to the formula for treating the regulatory capital value (RCV) in periodic price controls, something which would damage the incentive to invest. In a letter to managing directors in October 1995 (MD111) he had suggested that "simply adjusting the capital value to reflect actual investment with a five year lag is neither necessary nor sufficient.". The formula in place was that if a periodic review took place in 1999, planned net capex from 1995-1999 would be used to roll RCV forward, but actual net capex would be used for 1990-1994, replacing the planned net capex for 1990-1994. The uncertainty continued because a promised paper from Ofwat on the subject did not appear quickly.

Notwithstanding the uncertainty, Ofwat did show its capacity to impose a penalty on a company for failing to perform. Its 1995/6 enquiry into Yorkshire Water found "serious failures by the company in controlling leakage, minimising unplanned supply interruptions, and controlling flooding from sewers.". The DG agreed penalties with the company which included taking away the K factor for 1997/8 so that prices could only increase in line with inflation. It was a significant event because it was the first based on an agreed penalty, and Ofwat's press release said that "it is not the job of the regulator to manage the company" but where a company fails to perform "it will be required to undertake specific action, to report the outcome to the regulator and to be restricted in its ability to increase its prices.". The regulatory regime continued to be gradually toughened through to 1997.

To examine the main developments in water regulation from 1990 to 1997 in detail is best achieved from reading Ofwat publications throughout this period. At the end of 1990 Ofwat published a consultation paper on future charging policies for water and sewerage services entitled *Paying for Water – A Time for Decisions*. Whilst the price control formula could control the overall position (based on a 'basket' of tariffs, and the increase in average bills in real terms over the next ten years, resulting from the initial K values at privatisation, would range from 22% to 122%), it did not cover the balance between domestic and non-domestic users, or the economic usefulness of the tariff system adopted, given the majority of domestic customers paid in relation to the rateable value of the house they lived in. Ofwat was concerned that charging policies should achieve three elements:

- * Fairness and equity – to ensure customers in similar circumstances pay similar charges, and that charges were cost-reflective;
- * Sensible incentives – to help ensure that the right level of service is provided at the right price for customers and companies, and this would include charging systems which take full account of the cost of pollution. Ofwat noted that "Standing charges will give greater income certainty to the company than use-related charges. It would seem appropriate to reflect this reduced risk in the price limits set for companies with large standing charges." (p18). However, use-related charges would require an extension of domestic metering, but "studies in England and Wales have suggested that the impact of metering on demand is unlikely to be sufficiently great for the savings in resource costs to exceed the costs of immediate general installation of meters." (p22).

* Simplicity and comprehensibility – so that customers could be clear how their overall bill was made up, and what they could do to influence it.

These affect such things as the balance between standing and usage charges, and whether charges should be based on average costs or long run marginal costs. One paragraph is an important statement in support of the view that economic regulators are not ‘agents of social justice’. On page 10 it says “Difficult issues are raised by cases where water use is unavoidably high, perhaps for medical reasons (such as kidney dialysis), and by those where particular consumers have genuine difficulty in paying their bills. But it would be unfair to other water users if general tariff policy were to reflect social objectives. These should be part of health and social services policy. Any costs from providing support to customers with particular needs should be met by the appropriate agency, and not by water customers generally.”.

In order to avoid undue discrimination, companies had distinguished between costs of providing water and sewerage services, and between the costs attributable to measured and unmeasured customers. However, for Ofwat this raised the question on what basis was the attribution of costs made, particularly the allocation of common costs and financing costs. For Ofwat, this meant that companies should be required to disclose the basis of cost allocation in their regulatory accounts, and that that allocation should be expressed in current cost terms, given companies were obliged to produce current cost supplementary accounts to any main accounts which were on an historical cost basis. The cost of surface drainage makes up nearly half of the cost of the sewerage service, but it was thought that recovering surface and highway drainage costs through a banded standing charge might provide the best balance between fairness and simplicity, with foul drainage to be charged on a volume related basis.

Its overall conclusion was that “Charges for water and sewerage are not taxes, and so ability to pay would not be a good basis for them. Subject to avoiding complexity, charges should be based on costs, so that cross-subsidisation is avoided.” (p28). “The main options for future charging methods appear to be metering, banding of property, or a flat rate licence fee (essentially a 100% standing charge.” (p29). For domestic metering “It would be premature to take radical decisions while the meter trials are in progress and when public debate is at an early stage.” However, “In advance of, and perhaps in parallel with, any progressive extension of metering, it could be useful to develop methods of banding domestic properties according to estimates of the consumption typically associated with different size dwellings.” (p29).

1993 proved to be an important year for the development of the customer focus of economic regulation. At privatisation, Ofwat had ten regional Customer Service Committees but an Ofwat National Customer Council (ONCC) was established in March 1993 by the Director General of Water Services, with the support of the Secretary of State for the Environment and the President of the Board of Trade, to strengthen the representation of water customers’ interests nationally. The council’s members were the chairmen of each of the ten regional committees, the DG, and a chairman appointed by the DG. Its second annual report 1994-95 – The Voice of Water Customers makes particularly interesting reading because it covers the year in which Ofwat had completed its first periodic review of the price caps for 31 water companies.

The Chairman’s foreword (Jim Gardner) sets the tone. “At every stage of the periodic review process the Council was consulted by the Director General. This enabled the collective views of customers on a range of issues to be voiced directly to him. The Director General proved a ready listener, and a responsive one where there was room for manoeuvre in balancing his statutory duties. The way in which the Chairmen of the ten CSCs and the DG worked together during the Periodic Review was a significant milestone in regulation. The Council hopes that

there will be still more customer involvement in future price setting which will inevitably give more transparency to the whole process.” (p4). Given the Council’s view of the importance of affordability to customers, one view it expressed in the review of the Periodic Review was that “The Council would like to see a more public process including consideration given to publication of the companies’ ‘strategic business plans’ and the new price limits that they were seeking.” (p4). The Council’s concerns on customer bills were also well established in three areas. First, “Europe continues to concern the Council greatly. The massive investment necessary because of the Urban Waste Water Treatment Directive has been digested in the Periodic Review. But lurking around the corner are five more directives being processed by the European machinery which could have a major impact on water bills.”. The Council sought to influence the debate by meeting with staff of the European Commission’s Directorate General XI which was responsible for water quality and environmental matters, and kept in touch with the Environment section of the UK’s permanent representative in Brussels (p12). Second, on compensation for less than reasonable service, “I am extremely glad that, at long last, the water companies through their Associations have developed with Ofwat a framework document which lays the ground for developing an approach to compensation suitable for private companies who want to operate in a commercial environment. The Council, with the CSCs, will continue publicly to quote examples of good and bad practice in an endeavour to expedite the benefit to customers.” Finally, on paying for water the Council begun to question whether economic regulation could exclude social justice, saying “What is apparent is that increasingly some family units will continue to have difficulty in paying bills and that there will be continued pressure for a change in the law to enable companies to show (undue) preference to such customers in their tariffs. This illustrates extremely well just how political many regulatory decisions are seen to be – and why regulation will be to the forefront of debate until the general election (1997).” (p5).

One such political issue was executive remuneration. “Rightly or wrongly, rising water prices together with the controversy over executive pay has made the water industry the most unpopular utility. It is regrettable that the privatised water companies operating in a monopoly position have not recognised the need for more sensitive behaviour. The Council believes that the industry’s unpopularity has served only to get in the way of the real issues – the quality and efficiency of the service that it provides to customers.” (p10).

In terms of the development of the Council’s role, the annual report noted three features. First, Ofwat had modified its management structure in December 1993 and this involved the creation of an Ofwat Executive to assist in decision making and implementation. The Council was represented on the Executive (priorities) Group and the Executive (External Relations) Group. The Council concluded that it “is satisfied that these arrangements will ensure that it is, and will remain, at the heart of Ofwat which can only be to the benefit of water customers.” (p13). Second, it was “delighted to receive a reply from the DoE minister saying that the government shared the view that improvements in water environmental quality should be introduced at an affordable pace and that it was right that the Environment Agency should be under a duty to consider costs as well as benefits.” (p15). Finally, on the PR front, it was noted that the Council was represented on the National Industries Consumer Council’s (NICCs) Chairmen and Chief Officers’ Groups, which brought together representatives from the utilities as well as the Post Office and the railways, and was clearly pleased to find that the new journal Utility Week had carried a two page feature on the Council Chairman, with his photo on the front cover, entitled Customer Champion, in January 1995.” (p19).

By 1995, following the first five year price completed in 1994, Ofwat's Annual Report 1995 was able to record a number of 'firsts' given the five years since privatisation in 1989: the first price review, the first appeals to the MMC, the first inter-utility mergers, and the first time in which the companies' ability to provide a continuous supply of water was severely put to the test. Such developments had encouraged Ofwat to develop an 'output-based' regime in order to focus on the service that customers receive. Ofwat considered that "Perhaps the biggest challenge facing the regulator and the industry is to ensure that we have the public's confidence." This meant companies should deliver the outputs which customers are paying for, and Ofwat should "become increasingly open in providing information and explanations of the reasons behind regulatory decisions." In doing this Ofwat considered the customer representatives were vital: "In discharging my statutory duties on behalf of customers I will continue to rely heavily on the Ofwat National Customer Council and the ten Customer Service Committees. The Council has grown in stature during the year and is increasingly seen as an essential part of water regulation. The CSCs play an important role in monitoring company performance as well as investigating customer complaints and representing customer views." (The DG's Statement).

Two companies had not accepted Ofwat's price determination – South West Water and Portsmouth Water – and had appealed to the MMC. The MMC's judgment was to set price limits almost identical to those set by the Director General. Following guidelines issued in 1992 on disconnections procedures for customers that had not paid their water bills, Ofwat could report that in "1995-96, domestic disconnections at 5,826 are a quarter of the level in 1991-92 and well below the pre-privatisation level of 15,255 in 1988-89." The changing structure of the industry meant that the MMC was involved in reviewing proposed mergers and takeovers, covering the following:

- * Lyonnaise Europe Plc and Northumbrian Water – the MMC reported that the merger would be against the public interest, and the Secretary of State for Trade and Industry asked Ofwat to identify a solution. The result was that from 1996 customers would see price reductions which built up to 15% by 2001, while levels of service would be maintained.
- * Sutton District Water Plc and East Surrey Water Plc – as smaller companies the problem of a loss of a comparator was not so significant and they agreed to price reductions of 3.5% with the hope of a further 1.5% by 1999.
- * General Utilities Plc, Saur Plc and Mid Kent Holdings Plc – an automatic reference to the MMC given the turnover threshold.
- * Wessex Water Plc, Severn Trent Plc and South West Water Plc – both bids for South West Water Plc were referred to the MMC.
- * North West Water/Norweb and Welsh Water/Swalec mergers – These two cross-utility mergers were allowed to go ahead subject to licence amendments proposed by Ofwat and Offer to protect the interests of both water and electricity customers.

These developments meant that Ofwat reported on Competition in the Water Industry (inset appointments and their regulation) in 1995 and Transfer Pricing in the Water Industry in 1996. Although the DG only had a duty to facilitate competition, some competitive provisions were in place from the outset in that new suppliers could be allowed to operate, under licence, within the area of an existing supplier to provide services to a 'green field' site. This was known as an Inset appointment. Changes had been made by the Competition and Service (Utilities) Act in 1992 to extend the range of inset appointments that could be made (by including existing customers taking more than 250 megalitres of water per year), and the interest in this type of competition had increased, although Ofwat had only received three formal applications by 1995. However, the threat of such competition did persuade the water companies to introduce more cost-reflective tariffs for their larger customers, which meant Ofwat could report that 18 water

companies offered differential tariffs for high users of water, and South West Water included its sewerage services as well. Ofwat's report was aimed at making it clearer for those interested in an inset appointment, to simplify the procedures, and to ensure that whilst customers were protected they were not over-protected.

On transfer pricing water companies have a statutory duty to trade at 'arm's length' from their parent groups and prevent cross-subsidy to ensure that the costs of unregulated activities are not passed on to customers. The report had an important thing to say about the auditors of water companies. "It is not the regulator's job to audit company procedures. This process rightly remains the duty of the companies' auditors. But the reviews have shown the need for improved scrutiny of company compliance in this area by their auditors." The DG's foreword went on to say "Where we found non-compliance I have required companies to take corrective action. I am pleased that they have agreed to this. We will however keep up our checks to ensure that companies continue to follow the guidelines (issued in 1994). I do, however, have some concerns about a situation where trade with associates is a high proportion of the turnover of the regulated business and where the associates are heavily dependent upon trade with the regulated business. In such cases it could be better to reintegrate the activity into the regulated business."

By 1996, the MMC had blocked the merger of South West Water with either Wessex or Severn Trent, and the takeover of Mid Kent by Saur and General Utilities, both on the grounds of a loss of comparators. The 1996 Annual Report noted that a report prepared by London Economics on behalf of Ofwat had developed 'long run marginal cost' as a guiding principle for setting tariffs in a way which related to the continuing costs of augmenting supply. Ofwat concluded that the continuing cost of augmenting supply would generally be above the average cost of supplying existing volumes of water, something which would be particularly important for setting large user tariffs. One important issue to arise was that following the 1994 price review, a number of companies had said that the outputs could be delivered with substantially lower capital expenditure than had been allowed for in the price limits. This led to the following statement "The Director was concerned that rephasing capital programmes could benefit shareholders at the expense of customers. He therefore suggested to the companies concerned that they should not take up their full price limits in April 1997. Eleven agreed providing a total saving to customers of some £35 million."

The suggestion by Ofwat that water companies should voluntarily give up revenue within a price review period seemed to suggest that the accepted economic model of the RPI – x system was always going to be politically criticised, given the widely held belief that private profit was not appropriate for a 'public service' industry, and that economic regulators would have to take account of that. The forthcoming election in 1997 would also have provided some incentive to do so. An interesting seminar was held at the Peterhouse Theatre, Cambridge University, in March 1996, sponsored by Anglian Water. Entitled Public Acceptability and the Regulation of the Water Industry, it covered a wide range of issues, including the extent to which companies can develop systems of self-regulation and best practice to improve public accountability, and thereby reduce the extent of micro-regulation. Whilst the seminar had shown a general consensus on the need for incentive based regulation, the price and profit control system was an issue on which opinion was clearly divided. The water regulator, Ian Byatt, had suggested a 'steady as she goes' approach as being appropriate, but others wished to see the RPI – x system replaced by continuous profit sharing, although the response to this was that any attempt to direct profit sharing on a formulaic basis was likely to be manipulated. One important conclusion was that companies would have to address the issue of public acceptability through

the relationships that they are able to build with their customers. They could not look to the reform of regulation to cure current ills and public perceptions.

Ofwat contributed to the debate on RPI –x with a paper published in May 1997, entitled Profit Sharing. Its main conclusions were that:

- * On the method and timing of the transfer of benefits from shareholders to customers, the approach should be an initial downward adjustment of real prices in the first year of the new price limit period (often referred to as a Po cut methodology). If adopted then approximately 80% of the past outperformance benefits would be transferred to customers at each periodic review.
- * A shorter review period (say 4 years instead of 5) would undermine the incentives to outperform, as would formal profit sharing, and only result in customers having a larger slice of a smaller cake.
- * Companies which have voluntarily shared benefits with customers in the current price control period should be no worse off than those which have opted not to share at the next periodic review.
- * The benefits of profit sharing are more apparent than real, and would serve more to correct a public perception of water companies as making excess profits at the public expense than in increasing the public good by reducing prices to customers.

Ofwat's next annual report (1997) took account of the change of government in 1997, and is appropriate for the next period 1997 – 2004.

The Electricity Industry – privatisation and regulation 1990-1997

The electricity industry was prepared for privatisation in parallel with that for water privatisation but was sold in 1990 rather than 1989 for water. Electricity privatisation was the first significant change from the four previous privatisations in that the government restructured the industry prior to privatisation in order to provide a better platform for effective competition, and to respond to the criticisms received concerning the earlier privatisations that public monopolies had been privatised to create private sector monopolies. The government considered that electricity generation and the retail market could be competitive, leaving electricity transmission and distribution on an integrated network as a natural monopoly. This reflected the essential nature of electricity supply that demand must equal supply at all times, as in general surplus electricity cannot be stored, except by using the excess to pump water to high levels for hydro-electricity or to charge batteries.

The form of the industry which the government sort to privatise, and how it had developed, is comprehensively set out in the 1985 publication by the Electricity Council called Electricity Supply in the United Kingdom – A Chronology. In 1926 the Electricity Supply Act introduced the first effective national co-ordination by creating a public corporation, the Central Electricity Board, and the linking up of the existing regional system into a national 'grid'. The grid system (132,000V) was largely completed by the end of 1935 and by 1938 the proportion of spare generating plant had been reduced from 80% to about 15%, a significant reduction in costs. (45). In 1947 the Electricity Act brought the supply industry into public ownership, integrating 560 organisations into 14 new statutory electricity boards, twelve in England and Wales and two in southern Scotland, sharing responsibility for the retail distribution of electricity to consumers. Generation and transmission were the responsibility of the British Electricity Authority, and Electricity Consultative Councils were established for each of the 14 area boards. (60). The Electricity Act 1957 established two new statutory bodies, the Electricity Council for England and Wales and the Central Electricity Generating Board (CEGB). These two bodies remained

in place until privatisation in 1990. (73). The CEBG was reorganised in 1958 into five regions to allow some decentralisation of maintenance responsibilities. (76). In 1963 the UK parliament's select committee on nationalised industries reported on the electricity supply industry and made two interesting points, given the preparation for privatisation twenty years later, stating "the industry's structure appeared sound, and, if it was open to criticism, this should be confined to its performance." (86). The nationalised industries created a lot of academic and policy attention, and in 1968 Allen and Unwin published *Optimal Pricing and Investment in Electricity – An Essay in Applied Welfare Economics*, by Ralph Turvey. (97). In 1977 the Government replied to the select committee on nationalised industries report on Gas and Electricity Prices, accepting the need for a return to economic pricing, and a phasing out of subsidies to the nationalised energy industries, whilst measures were designed to protect the poorer consumer. (121). The Electricity Consumers' Council was formed in 1977, with statutory status granted by the Energy act 1983, and the chairmen of area electricity consultative councils were ex-officio members. (124). In 1981 the Hundred Group of Chartered Accountants published *The Financing of State Owned Industries*, arguing that the current system was rigid and inflexible and that it should be replaced by a form more nearly modelled on the debt/equity capital structures of private sector companies. (144). The government responded to the pressure to improve the performance of nationalised industries with a report in 1982 entitled *Accountability to Government*. (145).

In preparing for privatisation with a desire for some restructuring to promote competition the government was nevertheless limited by the economic realities of an industry where, unlike other industries, the requirement for instantaneous balance of generation and consumption had to be maintained. The industry had already developed a system of utilising generation plant in a cost effective way, where additional generating capacity is brought in on a 'merit order' to achieve the lowest overall cost for meeting demand. The merit order took account of the fact that different types of generating systems take different times to come up to full power, which meant there was a preplanned increase or reduction in generation to meet fluctuations in demand. This required close management and accurate demand forecasting throughout the day. This meant that nuclear stations, which have high capital costs but low fuel costs, are most economic when operated continuously and thereby met the 'base load' demand. Consumers electricity tariffs had been designed to encourage consumers to take their electricity at non-peak times (eg, night storage heating), and some flexibility was made available through 'Load Management Schemes' agreed between the CEBG and large electricity users.

The government's options were therefore restricted to restructuring the CEBG in such a way that the essential elements of the network system were maintained and consumers could be offered the chance to choose a different electricity supplier than their local area electricity board. The CEBG argued strongly against being restructured, particularly because of the advantage of co-ordinating the use of resources of the whole of the generating capacity. Investors expressed concern about buying the nuclear industry, given the high capital costs of new plants and the high cost of decommissioning old plant. The public were concerned about national security if there was to be foreign ownership, and also concerned about public safety if the electricity was privately managed. The fact that electricity prices were allowed to rise in preparation for privatisation added some cynicism to the political debate.

By December 1988 the government had made up its mind with the Electricity Bill being published. The CEBG would be broken up and consumers would be given the power to switch suppliers. The main features resulting from the abolition of the CEBG and the Electricity Council were:

- * A new company, National Power, would own 70% of the existing plant, including nuclear;
- * PowerGen would own the rest;
- * The twelve area electricity boards would become private distribution companies with an obligation to provide a supply at the request of a consumer;
- * The national grid would be transferred to a new company, jointly owned by the 12 distribution companies;
- * A Director General of Electricity Supply would be appointed to monitor the new companies and protect the consumer interest;
- * The financing requirements for the nuclear industry would be met from a 'non-fossil fuel obligation' imposed on the industry to ensure a diversity of energy sources, and the government would contribute to the decommissioning of nuclear power stations and the treating and storage of nuclear waste if the costs proved higher than forecast. However, as one commentator observed, "The Government has pushed ahead with its plans so quickly that the regulations and the contractual framework are having to be worked out with insufficient time for measured public discussion and the elimination of practical snags." (Public Domain, The Public Services Yearbook 1989, p97).

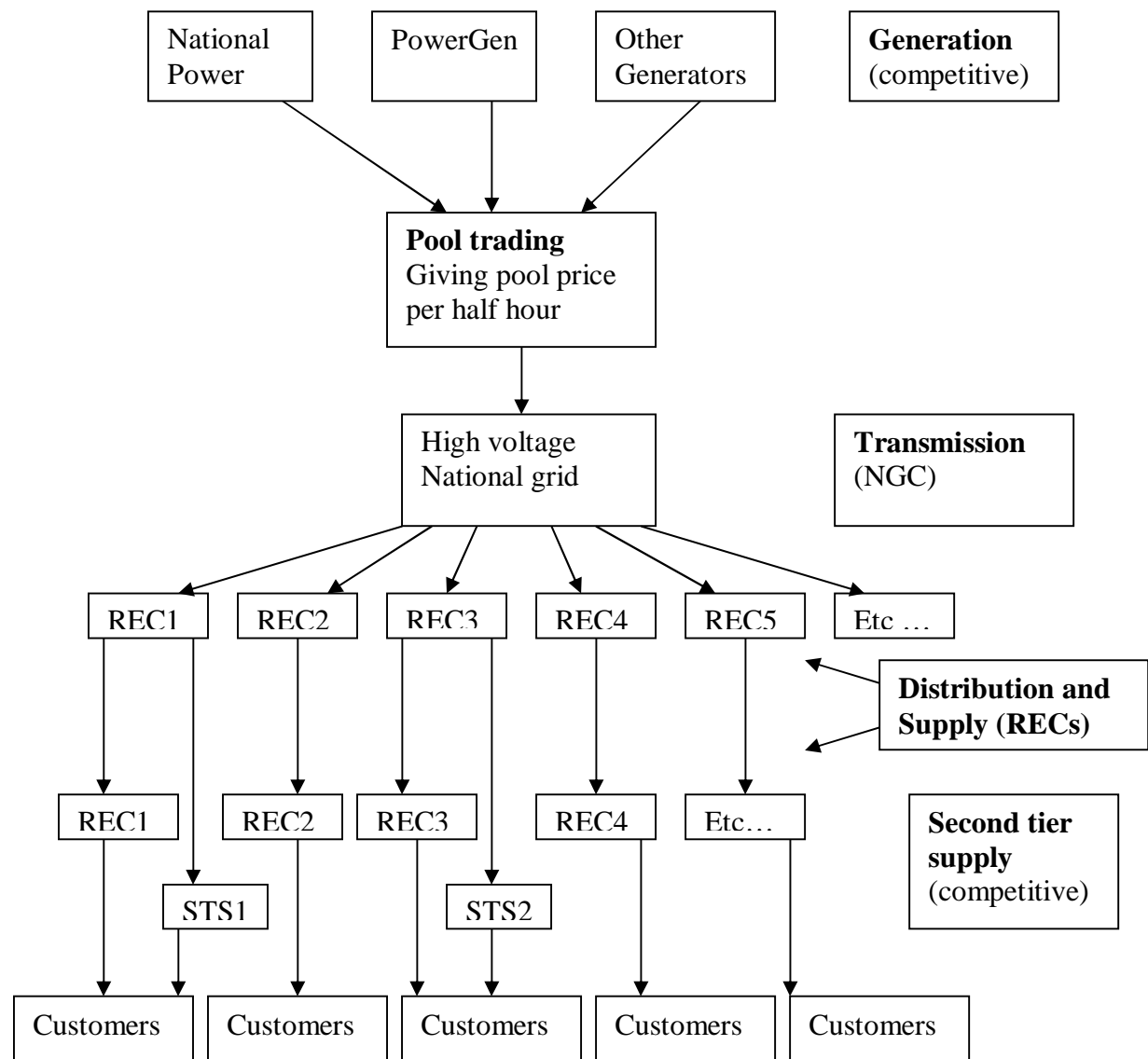
On the 1st September 1989 the Director General of Electricity Supply, Professor Stephen Littlechild, was appointed, and took responsibility for setting up the Office of Electricity Regulation (Offer) and its associated consumers' committees. The first annual report 1989 covered the four months since appointment and referred to the DG's main duties as:

- * Securing that all reasonable demands for electricity are satisfied;
- * Securing that licensees can finance their licensed activities;
- * Promoting competition in generation and supply.

The promotion of competition would in due course be revealed by the issue of new licences for new generators and suppliers, in part because transitional arrangements were put in place to allow a period of adjustment for the newly privatised companies. The annual report also focused on the hopes for developments in electronic metering whereby it would be practicable to determine in greater detail (by time of day and season) the quantities of electricity supplied to individual customers. The benefits of this would include pricing for supply which recognised the value of the call on available capacity, particularly the value of capacity at peak times. The Pooling and Settlement Agreement provided a means of incorporating into the pool price an administrative element intended to reflect the value of capacity and hence the value attached to meeting the security of supply against the generation security standard. A key element of this was the amount prescribed in the supply licences for the Value of Lost Load (VOLL).

The new arrangements for consumer affairs were the transfer of responsibilities from the area electricity consultative councils and the national electricity consumers' council to new 'Consumers Committees' within Offer. The DG did not simply stick to the existing area chairmen to appoint for the new consumer committees, stating "I wanted to signal the beginning of a new regime by introducing a significant number of new appointments....I therefore sought a balance between those with previous experience on consultative councils and those coming fresh to the subject." (p14). The first meeting of the 14 chairmen took place at the end of November 1989, chaired by the DG. It was agreed that this National Consumers' Consultative Committee would meet four times a year at various venues around the country. Offer also had responsibility for collecting the 'fossil fuel levy' which had been set by the government at 10.6% for the first year, a levy intended to ensure that the costs resulting from arrangements to meet the 'The Non Fossil Fuel Obligation' would be recouped from companies supplying electricity derived from fossil fuels.

The major restructuring of the electricity industry is shown in the following diagram. The diagram, with its arrows, shows why the customers electricity bill is made up of four parts: generation, transmission, distribution and supply.



By 1990 the government became increasingly aware that financiers and analysts in the City had serious doubts about privately-run nuclear stations, not only because of leaked reports from the CEGB that nuclear costs were far higher than expected, but also because of the danger of arrangements being ruled anti-competitive by the European Commission. In the end the Government decided to transfer all the nuclear assets to two new state-owned corporations, Nuclear Electric and Scottish Nuclear. Privatisation of electricity generation was now reduced

to two players, National Power and Powergen, and concerns about their potential abuse of monopoly power became more prevalent. The idea of a wholesale market in electricity was also effectively scrapped, as the plan for a network of separate supply contracts between power stations and distribution companies was found to be difficult to operate effectively. The new National Grid company reverted to the pooling of outputs from all the generating companies and the setting of a single price per unit of electricity, which would change every half-hour depending on supply and demand. This was the system broadly used under nationalisation, and reflects the reality that you cannot link a particular unit of consumption with a unit of supply.

The 12 regional electricity companies (RECs) were sold in December 1990 (with each REC owning between 5.4% and 12.5% of the shares of National Grid holding plc), and National Power and Powergen were sold in 1991. The RECs distribute and supply electricity to domestic, commercial and industrial customers, but can also retail electrical appliances, provide electrical contracting services, and, in some cases, generate electricity. The new freedom for customers to choose their supplier also meant that not every user to whom a REC distributes electricity will be a supply customer. As the DG said in his first annual report “When companies know that customers have freedom of choice, they try even harder to deliver a service that meets consumers’ wishes.” (p5). The sale was massively oversubscribed (the offer price of 240p per £1 share) and in January 1991 the shares were trading at premiums of between 17% and 27% of the full offer price. Not surprisingly, given such premiums represent close to £1000m, there was some criticism that the proceeds of sale of £5182m could have been significantly higher.

In terms of the new regulatory system for electricity price control, there was no requirement for direct regulation of generators prices. The NGC had responsibility for bringing in generators to meet demand on the basis of a merit order of prices and availability, and those generators offering the lowest prices would generally be used first. The DG would have to be aware of any cartels developing between generators, but in general the price of generated electricity was expected to rise or fall over time depending on the market conditions and the technological factors which underpin the real costs. Overall the pool price of electricity was established every half-hour and the components of the price include a marginal cost element, a capacity element, and an uplift to reflect in part the requirement for a reserve of generators and the maintenance of a stable, integrated system. Price regulation was therefore applied to transmission, distribution and supply, with each REC required to run distribution and supply as separate businesses. The initial formulae were:

Transmission (NGC)	$RPI + X_t$
Distribution (RECs)	$RPI + X_d$
Supply	$RPI + X_s + Y$

with X_t and X_s initially being set a zero (which effectively holds prices constant in real terms), and X_d ranging from 0 to +2.5, which reflected the need for real increases in prices to finance different capital expenditure requirements. The Y term in the supply formula allowed suppliers to pass on costs to customers where these were:

- * Electricity purchase costs, including direct pool purchase costs and associated net costs of ‘contracts for differences’;
- * Transmission charges from NGC;
- * Distribution charges from RECs;
- * Settlement system charges;
- * Fossil fuel levy.

The risks to the newly privatised RECs were abated in two ways. First, they were given 'transitional' protection, which meant the RECs had the sole right (or franchise) to supply almost all premises in their areas until 1998 (sales less than 1mw to march 1994 and 100kw to march 1998). Thereafter, all customers would be able to contract a 'second tier' supplier. Secondly, because RECs supply electricity on fixed tariffs and contracts, but face exposure to variable pool prices for purchasing generated electricity, the RECs have sought to limit their exposure to pool prices by entering into 'contracts for differences' with the generators, and thereby spreading their risks. A fee is payable to the generators in return for payments by them to the suppliers when the pool prices exceed certain levels. These contracts give greater income certainty to generators and greater stability to suppliers in setting tariffs.

Offer's first full year annual report (1990) drew attention to how Offer should respond to developments as privatisation progressed, and was firmly positive in outlook. Interesting statements by the DG included: "Some people were apprehensive about the arrangements that had been designed to operate in this new regime. Other people, both abroad and in the European Commission, are actively considering new policies which might incorporate some features of this regime. Observers are therefore looking with particular interest at developments in Britain." (p2). "There are also encouraging signs of new attitudes and thinking throughout the electricity industry. The most visible manifestations of this are the decisions by generators concerning the construction of new plant. In all the firms I have talked to, I have been impressed by the systematic and far-reaching reappraisals of policy, notably with respect to investment, operating procedures, staffing and organisational structure. Their aim is to operate more efficiently and entrepreneurially while maintaining or improving the standards of service offered to customers. I believe this rethinking reflects the combined effects of the incentives of private ownership, the pressures and opportunities of competition, and the framework imposed by the new regulatory regime." (p3). However, Offer did not rely on the new private electricity suppliers to maintain or improve the quality of service, as it could set standards of performance which must be met by the PESs. It said "My general aim has been to set standards which are realistic but challenging. I have therefore consulted Consumers' Committees and other consumer and industrial groups as well as the companies. Detailed plans for the new standards were well advanced by the end of the year (to be supported by Codes of Practice and with compensation payments were appropriate). I expect that most of the standards will come into operation by the middle of 1991." (p18). Looking ahead, one statement was the reinforcement of the intention to issue new second tier supply licences to all credible suppliers in order to encourage competitors to the RECs in their own areas, and this was reinforced by the recognition that the subsidiary price caps on franchise customers might be relaxed if, in the new and uncertain market, costs increase for RECs through unavoidable circumstances. The latter reflected the obligation on the DG to ensure that companies can finance their operations adequately.

Three reports were published in 1992 by the National Audit Office on the sale of various parts of the UK electricity industry: The Sale of the Twelve Regional Electricity Companies (14th May), The Sale of National Power and PowerGen (11th June), and The Sale of ScottishPower and Hydro-Electric (15th July). In the first report its analysis of the government's success in meeting its main objectives (complete the sale in the lifetime of the Parliament, maximise net proceeds, widen and deepen share ownership, overall recognition of success, and achieve a modest premium over the issue price of 240p) came to an important conclusion on the method and timing of the sale. The premium at which the shares were traded when dealings began averaged 49p compared with the government's target of 15p. Of the 49p premium some 30p was attributable to the general upward movement of the market which began after the offer was

priced, but this still left 19p which was in excess of the target 15p. The general conclusion reached by the NAO was that “A number of previous privatisations had proceeded on the basis of a partial sale. Having considered whether they should follow this course, which all their main financial advisers strongly recommended against, the Department for Energy went ahead with a 100% sale. This was on the grounds that not to have done so might have jeopardised the success of the sale and the government’s overriding objective of completing all three electricity sales to a demanding timetable. Nevertheless, the novelty of the industry, and of the companies, and the uncertain market conditions at the time of the sale, arising out of the deepening recession, the Gulf crisis, and the Conservative leadership election, made it particularly difficult for the Department to price the issue.” (p6). The NAO’s recommendation was that departments should continue to give careful consideration to retaining a substantial minority shareholding to be sold at a later date, when advantage could be taken of the further information which would emerge from a longer track record.

In the second report, on the sale of National Power and PowerGen, the NAO noted that the Department had decided only to sell some 60% of the shares, and had a more innovative offer structure. A proportion of shares was provisionally earmarked for offer to the highest bidders from institutional and overseas investors after the close of the offer period (back-end tender); institutional and overseas investors were required to indicate, in advance of the pricing of the offer, how much they would be prepared to invest at a range of prices (bookbuilding); and the department dispensed with primary underwriting. The NAO could conclude “In the round, the innovations adopted by the Department brought worthwhile gains to the taxpayer” and recommend for future sales that “wherever possible continue to include in the team staff experienced in previous sales.” (p5).

The sale of the two Scottish electricity companies was the last stage in the government’s programme to privatise the electricity industry in Great Britain, and reverted to selling 100% of the shares. The reason for this was that the sale was taking place in a sector now established in the market, the pricing techniques had been successfully applied in the previous electricity sale, and in favourable market conditions. The NAO concluded that the sale had been on favourable terms for the taxpayer, and had secured proceeds which were as high as could have been expected. In terms of the objective to widen and deepen share ownership, the NAO could say “The Department met their target for subscription levels from individual investors and were able to allocate a large proportion of the issue to individual investors, particularly to customers, employees and pensioners of the two companies. Overall, share retention levels have been higher than in the previous two electricity sales but there have been significant disposals by those individuals who were allocated the most shares.” (p5).

In addition to selling the UK electricity industry, the government had announced in 1988 its intention to privatise the electricity supply industry in Northern Ireland, run by the public corporation, Northern Ireland Electricity (NIE). The privatisation took place in 1992 for NIE’s four generating stations which were sold to three private companies by ‘trade sales’, and in 1993 for the remainder which was floated on the London stock exchange. The sales were carried out on the government’s behalf by the Northern Ireland Department for Economic Development. The Committee of Public Accounts reported on the Privatisation of Northern Ireland Electricity in May 1995. It was critical of both sales, stating that “We do not agree with the Department’s view that its valuation of the generating stations was based on the best information available at the time, and that a proper valuation was made before a buyer was sought. We do not consider that it is sufficient to depend on competition between a number of bidders to produce the best price.” On the flotation of NIE it stated “The evidence strongly suggests to us that the shares in

NIE plc were undervalued by the department with the result that the taxpayer received less from the flotation than ought to have been the case.” (para 5). It was also concerned about conflicts of interest in that “One of the department’s advisers in the sale of the generating stations also provided advice during the tendering process to one of the eventually successful bidders.” (para 12). The PAC considered therefore that the taxpayer had been denied maximum receipts and stated “the Department must bear the responsibility for this but we consider that in some critical areas it was not well served by its advisers.” (para31).

In December 1991 Offer published its Report on the Pool Price Inquiry which had arisen because of concern over the manipulation of pool prices. Generators receive the Pool Purchase Price (PPP) plus an ‘uplift’ to reflect the costs of operating a safe and stable network, the total being the pool supply price. The PPP is made up of two components, the system marginal price (SMP) and an available capacity component. The SMP is derived from the offer prices for the marginal generating set (usually the most expensive) scheduled for the relevant half an hour, and the value of the capacity component depends on the probability of losing load and the value of the lost load. Offer’s conclusion was that “the dominant market position of the two generators gives them the ability to influence and control prices”. To promote competition a new licence condition was proposed specifically to prohibit monopolistic or anti-competitive behaviour in relation to the availability of plant and the closure or mothballing of stations. The licence change would also require arrangements to establish whether others would be willing to purchase any power station the generators intended to close or mothball. By 1992 a number of the regional electricity companies were investing in new gas fired generation plant (the so-called ‘dash for gas’) and Offer concluded that these were in accordance with economic purchasing obligations, and not simply monopolistic actions which resulted from the ability to pass through gas purchase costs to the consumer (Review of Economic Purchasing, Interim Report, December 1992). In any event, as second tier supply competition developed with the reduction of the franchise monopoly by 1998, cost pass-through of generation costs would be no protection to regional electricity companies that could not find customers to supply electricity to because they had invested in uneconomic generation. However, in 1993 Offer announced that it would continue to monitor the pool price and discuss with generating companies their plans to dispose of some of their plant to independent owners in order to assist competition. In addition Offer would make a decision on whether to refer the generators to the MMC by no later than 1995. The pressure to act was sufficient however to cause an agreement to be reached on 12th February 1994 which avoided a reference to the MMC but introduced a price control on generation costs in England and Wales for the first time. It was agreed that National Power and PowerGen would dispose of 600Mw (about 15%) of their generating capacity.

The economic regulators take the opportunity of the periodic price cap reviews to examine both the principles and practice of regulation, and many were lessons in economics and accounting. Overall there was a progressive tightening of the price caps at each review, in part due to passing on incentive efficiency improvements, and part any generosity in the price caps at the time of privatisation. In July 1992 Offer published Future Control on National Grid Company Prices – A Statement which expected a further 2% per annum reduction in operating costs on top of the 3% per annum proposed by the NGC in its forward plans. Offer also argued that the allowed cost of capital could be reduced because NGC’s business is significantly less risky than most other companies. Offer revised the price cap from RPI – 0 to RPI – 3 for a four year period from 1st April 1993. The supply price control review was also started in 1992 (the Consultation Paper published in October 1992), to be concluded in 1993. A key issue for the supply price review was whether the overall price control from 1993 should be limited only to the franchise market where competition did not exist. Offer’s preliminary conclusion was that it should,

subject to technical developments in metering and acceptable methods for apportioning generating costs between the two markets. The four year price control period chosen was judged by regulators to be an appropriate balance between the desire not to erode the incentive properties of the price cap, and the difficulties caused by uncertainties in the forecasting.

In 1993 the supply price review was concluded, but it is important to note that it was only concerned with about 5% of the bill for electricity because the other stages in the supply chain (distribution, transmission and generation) were dealt with separately and passed through to customers in the +Y factor in the supply price formula. The supply price controls were tightened from the existing $RPI - 0$ for all the twelve RECs to $RPI - 2$, so that real electricity prices would be reduced in real terms for the period 1st April 1994 to 31st March 1998, which coincides with the ending of the monopoly franchise. One interesting thing to emerge from the transmission price review for Scotland arose because both Scottish companies had plans to enter the telecommunications market, including offering services via telecommunications cables run along their transmission networks. To benefit electricity consumers, Offer netted off from the operating costs of each company an amount representing the telecommunications rental which the transmission business might be expected to receive.

The privatisation price caps for electricity distribution reached the end of their first period on 31 March 1995, so Offer began its review for the next period with a consultation paper published in 1993 entitled Electricity Distribution: Price Control, Reliability and Customer Service. This was a significant review because the main profits for the RECs come from distribution rather than supply (distribution charges making up some 22% of the final price for electricity). Given the concern over the rapidly rising profits of the RECs, the consultation paper focused on:

- * The structure of the distribution price control: this arose because although the formula constrained the average revenue per unit distributed, the base price was made up of a weighted basket of four component prices which differ between companies depending on each company's mix of customers connected to the system (three low voltage categories of peak, off-peak and standard domestic, and one high voltage). The question was whether to constrain the rebalancing which can take place within the formula, and how the structure of charges should relate to the attributable costs of each of the categories. Account also had to be taken of the distribution losses adjustment in the formula to ensure that it would be worthwhile for companies to buy low loss rather than standard transformers.

- * Metering: this arose because the franchise limit was reduced to 100kw peak demand in 1994, which increased the number of potential second tier customers from about 5000 to 55,000, and would be extended to all customers in 1998. Metering is an essential component in making supply competition possible over a wide range of small customers, but sophisticated meters are expensive and therefore limit the scope for effective competition. The question was whether to allow for sufficient investment by the RECs to arrange for a universal provision of metering which would meet the needs of competing suppliers.

- * Other issues, such as connection charges and standing charges.

By 1995 interest in takeovers and mergers was increasing because the government's controlling shares lapsed in 1995, and the government announced its intention to sell the remaining 40% of its shares in National Power and PowerGen. Trafalgar House made a bid for Northern Electric. The National Grid Company (NGC) owned by the 12 RECs was planned to be sold in 1995 because of the problem of NGC getting a clear direction from its 12 owners who were progressively competing against one another and therefore having different interests in how the NGC should operate and perform. The sale of the NGC raised one political problem which would arise if the RECs received much higher proceeds than the value which had been placed on

the NGC at privatisation – in other words, the government had sold the assets too cheaply. The distribution price review was completed in 1994 to set new price caps to run from April 1995. Compared to the privatisation price control period which had allowed prices to rise by up to 2.5% pa in real terms, the 1995 review resulted in sharp reductions of -11%, -14% or -17% depending on the company. The intention of this was to bring the companies back to a real reduction in prices relative to the tariffs at privatisation. For the following years the X factor would be – 2% pa until 2000.

However, the price caps agreed for the RECs did not last long as Offer announced in 1995 that it was reopening the price review on the distribution businesses in England and Wales. This was in part due to the fact that following the announcement the rising share prices indicated that the price cuts had not been unduly harsh. Most important though was that Trafalgar House's for Northern Electric in 1994 had revealed new information about the potential efficiency of the company, arising from the management's defence against the bid. Professor Littlechild, the Director General, took some time to decide to take this new information into account and reopen the periodic review because it raised serious questions about the credibility of the regulatory system. The credibility of the 'incentive' regulatory system is dependent on due process, a firm announcement, and the ability of the companies to be free to work within the price control for the agreed period. Nevertheless, the review was reopened, but in the context that steps would be taken to ensure that it did not happen again. The revised proposals were to introduce a further one-off cut in 1996/7 of, on average, 11.5% and a further tightening of the annual X factor from 2% to 3% pa until 1999/2000. It is interesting to note that none of the 12 RECs appealed to the MMC against the dramatic redeterminations.

The removal of the government's special shares in 1995 meant that takeover activity increased significantly, and the competition in the capital markets showed that even natural monopoly businesses are subject to competitive disciplines on their management. Takeovers were both domestic and international, with Scottish Power successfully bidding for Manweb and US companies (Southern Electric International and CSW) taking over SWEB and Seeboard respectively. The Hanson conglomerate acquired Eastern Electricity. North West Water created the first joint-utility company by taking over Norweb, creating United Utilities, and Welsh Water acquired SWALEC. The main regulatory response to such takeovers was to ensure that each utility business should be 'ring-fenced' to protect consumers and minimise the risk of cross subsidy. The proposed acquisitions of RECs by the major generators (National Power for Southern Electric and PowerGen for Midlands Electricity) raised concerns over vertical reintegration and an increase in market power, and these bids were referred to the MMC.

In the year before the change of Government to New Labour, 1996, British Energy was sold, achieving what had not been possible in 1991, which was to privatise the nuclear generators, although the Magnox stations which were nearing the end of their generating life and had substantial, and unknown, decommissioning costs were retained in public ownership. Northern Ireland Electricity faced its first periodic review, and NGC faced its second periodic review. After four consultation papers, the Director General made a tough settlement for NGC, cutting X from -3% to -20% in 1997 and then -4% pa until 2001. The decision of NGC not to appeal to the MMC was rationalised in terms that an appeal would tie up too much senior management time which was needed to focus on developing its business overseas and in telecoms through Energis, its company which could take advantage of the national grid lines to carry new fibre optic cables. The proposed cuts for NIE, which were for X to be -31% and then -2% until 2002, were appealed and referred to the MMC.

The restructuring of the electricity industry at privatisation meant that the regulator had many more periodic price reviews to carry out than the regulators for the previous privatisations, and the following table shows the price cap formulae which had been put in place up to 1997.

Transmission

NGC (England and Wales)		1990-93	RPI – 0
		1993-97	RPI – 3
		1997-98	RPI – 20
		1998-2001	RPI - 4
Scotland	Hydro	1990-94	RPI – 0.5
		1994-99	RPI – 1.5
	Scottish Power	1990-94	RPI – 1
		1994-99	RPI – 1

Distribution

RECs (England and Wales)	1990-95	RPI + 1.3(av)	range 0 to +2.5
	1995-96	RPI – 14(av)	range -11 to -17
	1996-97	RPI – 11.5(av)	range -10 to -14
	1997-2000	RPI – 3	
Scotland	Hydro	1990-95	RPI – 0.3
		1995-2000	RPI – 1
	Scottish Power	1990-95	RPI – 0.5
		1995-2000	RPI - 2

Supply

RECs (England and Wales)		1990-94	RPI – 0 + Y
		1994-98	RPI – 2 + Y
Scotland	Hydro	1990-95	RPI – 0 + Y
		1995-98	RPI – 2 + Y
	Scottish Power	1990-95	RPI – 0.3 + Y
		1995-98	RPI – 2 + Y

Northern Ireland

Transmission and distribution 1992-97			
	fixed component (75%)	RPI + 3.5	
	variable component (25%)	RPI + 1	
Supply		RPI + 0 + Y	

British rail privatisation and regulation to 1997

The privatisation of British Rail was the last major privatisation of the Conservative administration before the change in government in 1997. In some respects this privatisation was the most interesting because it was the most complicated, building on the restructuring approach of privatising the electricity industry, but having to cope with introducing competition by a different means to reflect the different operating characteristics of railways compared with electricity. These differences also introduced new regulatory bodies so that passenger rail

services could be dealt with differently from freight services, and account had also be taken of the fact that the railways were the only utility industry which received a large public subsidy. Rail privatisation also introduced a range of political problems not faced by the other utility privatisations, and it is useful therefore to start with the state of transport policy before the Conservative administration came to power in 1979. This is best reflected in the white paper published by the Labour administration in 1977, entitled Transport Policy (cmd 6836).

The purpose of the white paper was to achieve an 'integrated plan' for transport policy covering all means of transport, from aviation, road, rail, buses, lorries, and bikes and pedestrians. The objectives were set out in the first section and includes the principle objectives of transport policy which were (para 19):

First, to contribute to economic growth and higher national prosperity;

Second, to meet social needs by securing a reasonable level of personal mobility, in particular by maintaining public transport for the many people who do not have the effective choice of travelling by car;

Third, to minimise the harmful effects, in loss of life and damage to the environment, that are the direct physical result of the transport that we use.

In addition, the transport policies should support the national aims of energy conservation (para 30) and in doing that the policies for transport and energy must be linked with planning policies so that it fits in with plans for housing, employment and other uses of land, and the patterns of where people live and work, in conurbations or new towns, inner city area or estates on the outskirts of towns, or in the country (para 33). Taken together, it is no surprise that the development and regulation of the rail industry would be difficult.

The government recognised that these policies could not be achieved by centralised national planning, and that local authorities should have more responsibility for dealing with local transport problems, particularly as the real problems of co-ordination are local. Do the bus and rail timetables match? Are the bus stops well placed? And the car parks? What is the right balance between the free movement of traffic and the provision of public transport? The role of central government was therefore to formulate the national policies and provide the instruments of regulation, finance and taxation which would guide the various transport sectors to a balanced outcome. The question of allocating public expenditure between road and rail was particularly controversial in terms of the fairness of competition between rail and road. Many road users believed that they paid more in transport taxes than their share of the road costs, and therefore building more roads would be fairer than subsidising British Rail. However, the external and environmental costs should also be taken into account.

The white paper stated that the government was committed to maintaining adequate systems of public transport and that this would require substantial and continuing subsidy (para 55), but also said that subsidies must be controlled and not open-ended, and that growth in subsidies should not in effect be financed by a reduction in investment because of the overall limit on public expenditure. One implication of this was that transport operators would have to design their fare structures appropriately to reflect the capacity of the network, particularly peak and off-peak, and the associated costs that go with it, given the much lower resource costs associated with off-peak travel. The other implication of this for the railways was whether some railway lines should be closed. The government was clear that the rail system was required for long distance passenger services, and intercity services, transport services for heavy flows of bulk freight, and important commuter services into the major cities. However in local services there were some which carried few passengers at high cost, and were not necessarily a convenient way of meeting the travel needs. The government rejected the notion that that this problem should

be handled by the imposition of major cuts to the railway system, and considered that the best way forward would be to introduce a new procedure and responsibilities for considering line closures (para 99). The British Railways Board would identify the services which would not be worth retaining unless there was a strong local need for them. In doing this it would apply a value for money test. If the Board wished to close a line it would inform the secretary of state for transport. Following consultation the secretary of state would decide whether the line should be retained, and if it should the government would continue to support it and the Board would be obliged to run it under the national Public Service Obligation. An alternative would be to ask the local authority whether it wished to support the line, and if not what alternative service could be provided by bus operators, which the local authority would be responsible for contracting to meet the local travel needs. For the London and South East rail services the government decided not to impose on the BRB a specific financial objective for reducing the subsidy, but expected the Board, with the unions, to intensify their efforts to reduce costs and to phase in fare rises so that commuters had a period of years to adjust to them.

1997-2008 New Labour

Contrary to what many people had hoped New Labour would do, which was to renationalise some of the industries which had been privatised, they focused on consumer protection, deregulation and accountability for the economic regulators. The previous government had had a Deregulation Unit but this was renamed the Better Regulation Task Force (BRTF). Its remit was to “advise the government on action which improves the effectiveness and credibility of government regulation by ensuring that it is necessary, fair and affordable, and simple to understand and administer, taking particular account of small business and ordinary people”. One of their first tasks was to set out the Five Principles of Good Regulation, which were:

- * Transparency
- * Accountability
- * Targeting
- * Consistency
- * Proportionality

Accompanying this was the requirement on all government departments and economic regulators to carry out a Regulatory Impact Assessment (RIAs), both before and after a regulation was introduced. The RIA was based in the Cabinet Office under the Modernising Government programme. The relationship between policy and regulation was highlighted in its publication *Good Policy Making: A Guide to Regulatory Impact Assessment*. A further change was introduced after a debate on whether there should be a DG or a board. The BRTF said in its report on economic regulators that “Our final question to shareholders was whether different corporate structures influence regulatory effectiveness. One response was that this was yesterday’s question – the consensus is now that a board is preferable to an individual”. The BRTF also increased the emphasis on representative groups involved in consumer affairs. The question of standards to be imposed on regulated companies was an important part of this. By 2005 the BRTF was publishing *Routes to Better Regulation: A Guide to Alternatives to Classic Regulation*; *Better Regulation: From Design to Delivery*; and *Regulation – Less is More, Reducing Burdens, Improving Outcomes*.

The increased emphasis on the social responsibilities of the economic regulators soon found its way into successive publications from their offices. A good example is found in the new Office of Gas and Electricity Markets (Ofgem), which was established early in 1999. In October 1999 it published *Social Action Plan – A Framework Document*; in January 2000 it published *Social Action Plan: Enhancing Social Obligations, A Proposals Document*; and by March 2003 it published *Social Action Plan – Annual review 2003*. The purpose of these documents was to

“advance the process of identifying: what are the causes of fuel poverty, who affects it, the various bodies and organisations which can contribute to reducing or eliminating it, and what the special contribution of regulation should be”. These bodies would include government, local authorities, providers of banking services, voluntary organisations and consumer groups, as well as the suppliers of electricity and gas. The Electricity Association set up a Fuel Poverty Task Force in May 1999. Clearly under New Labour the economic regulators were to be agents of social justice. This arose as part of government policy set out in the March 1998 Green Paper, *A Fair Deal for Consumers*. An important part of the process was that Ofgem proposed revised licence obligations which would establish a common regime for electricity and gas companies on social obligations, including payment methods, energy efficiency advice, services for vulnerable customers and debt and disconnection. In March 2003 the government published its first annual review of the fuel poverty strategy which showed that since 1996 the number of households in fuel poverty (meaning that more than 10% of their household income had to be spent to heat the home adequately) had been reduced from five and half million to three million. Suppliers continued to develop innovative schemes to benefit low income customers. Through easy payment schemes, competitive tariffs, and partnerships to deliver energy efficiency measures, Ofgem considered that suppliers had responded positively to the social agenda.

The National Audit Office published an overall report on Pipes and Wires in April 2002, focusing more on the natural monopoly elements of the supply chain. This covered the successes and potential limitations of the RPI – x process, and particularly the strength of the incentives, distortions related to investment in networks, insufficient or excessive returns to investors, perceptions that the regulatory regimes are uncertain, and that the process of price regulation could impose excessive demands on companies. Its conclusions and recommendations were that:

- * The regulators should seek to remove features of their price reviews that give companies an incentive to bias their decision making or accounting to obtain more favourable treatment.
- * The regulators should consider publicly identifying the improvements in outputs and outcomes that they are willing to allow companies to invest in.
- * The regulators should encourage network companies to develop risk management models to assess the potential impact of deterioration in asset performance on future levels of service.
- * The cost of capital allowance should be set in such a way that transparently and consistently reflects the returns investors expect from investing in the companies concerned, assuming that they are efficiently operated.
- * The regulators should present clearly the assumptions and financial models underlying their price review decisions and the extent to which these will apply at the next price review.
- * The regulators should clearly specify and well in advance what information they will need from companies during price reviews and gather as much of that information as is cost justified on an annual basis.
- * The regulators should publish an evaluation of their completed price reviews, and in doing so should evaluate the different types of analysis undertaken to determine whether some would not in future justify the cost to both regulators and companies.

Developments like this help to explain why, over time, the length of the price review process increased, as did the number of publications from regulatory offices.

Apart from the development of accountability, better regulation and social action plans under New Labour, there were two instances where the basic model of privatisation which had remained in place through the Conservative and New Labour governments, and the Conservative/Liberal Democrat coalition when Royal Mail was privatised, was changed. One involved Welsh Water, which had acquired a regional electricity network and had branched into

other businesses such as hotels. It gradually became insolvent and was restructured to remove shareholders and become a not-for-dividend commercial company, limited by guarantee, and financed by debt. The other was Railtrack between 2000 and 2002, which the government put into administration following a huge increase costs resulting from a number of train crashes and derailments which occurred, in part due to the contractors appointed by Railtrack to maintain the rail network. It too became a not-for-dividend company, Network Rail.

At the time of Railtrack's flotation the prospectus contained a statement by the rail regulator. This included his expectation that Railtrack would develop, and deliver on, plans to maintain and renew the network in modern equivalent form, and take a proactive role in enhancing the network. By the end of 1996 it was clear to the regulator that it had not increased the rate of spending necessary to deal with the accumulated backlog, which he found totally unacceptable because Railtrack was not constrained by the external financing limits which were in place before it was privatised. The next step begun to raise questions about the independence of the economic regulator following a report in March 1998 from the Select Committee on the Environment, Transport and Regional Affairs entitled *The Proposed Strategic Rail Authority (SRA) and Rail Regulation*. The Committee had identified some overlapping functions between the rail regulator and Opra, and recommended the setting up of the SRA which would be accountable to the secretary of state for strategic planning, co-ordinating and supervising the activities of the rail industry, and for the disbursement of public funds. The Committee proposed the transfer of a number of functions from ORR to the SRA, with the rail regulator having much more limited functions of "umpiring contracts and fair competition issues". The government confirmed this with the Transport White Paper in July 1998 and *A New Deal for the Railways* (Cm4024). The increasing political involvement by government was highlighted by the secretary of state issuing new Objectives Instructions and Guidance to the Franchising Director. The franchising director had to take account of wider government policy and the moves to integrated transport. At the same time the relationship between the New Labour Government and the Rail Regulator was reflected in a 'Concordat'. Although the concordat confirmed the independence of the of the rail regulator, the government's aims and duties are important in interpreting his statutory under the Railways Act 1993.

The Competition Act 1998 in fact gave new powers to the rail regulator because the act prohibits agreements which have as their object or effect the prevention, restriction or distortion of competition, and outlawed the abuse of a dominant position. However, the Transport Act 2000 transferred the consumer protection functions of the regulator to the SRA, which included licence conditions relating to fares, insurance, policing and security, timetable information, enquiries, sale of tickets, through ticketing, disabled persons, and liaison with the Rail Passengers Council and the regional Rail Passengers Committees. The Transport Act also gave the regulator new duties to take account of guidance from the secretary of state and to facilitate the furtherance by the SRA of any strategies which it formulated with respect to its purposes. Nevertheless, the new rail regulator, Tom Winsor, emphasised that the secretary of state, the SRA and the regulator were all committed to the principle and the reality of 'joined' up regulation.

However, the Hatfield derailment in October 2000 changed the context and by October 2002 Railtrack had been placed into administration and its assets transferred to Network Rail, the second company to be limited by guarantee. The Hatfield derailment had led to the imposition of over 1000 temporary speed restrictions across the network, with its consequential widespread timetable disruption, and an unprecedented programme of rail renewal to address the phenomenon of rolling contact fatigue and gauge corner cracking in rails. The cost of this was

estimated at £600m. The rail regulator issued a policy statement on interim price reviews if change of circumstances was sufficient to make Railtrack insolvent. The government, the SRA and Railtrack also embarked on an exercise to consider reprofiling revenue grants to improve Railtrack's cash position. The regulator felt that the company had compromised its position as an independently regulated business by its direct approaches to government on funding issues, saying in a speech that "Railtrack should put away the begging bowl, and stop spending valuable management time hawking themselves unwanted around Whitehall, and knuckle down to getting train services back to a sustainable level of reliability and quality of service". By 2001 the secretary of state Stephen Byers successfully petitioned the high court to place Railtrack into railway administration. Although Railtrack continued to ask for financial support, Byers said that government could not justify any additional public money for Railtrack because "Railtrack Plc was taken into administration because it was, or was likely to become, unable to pay its debt, and with no shareholders we would remove the conflict between the need to increase shareholder value and the interests of rail passengers".

During the period of administration Network Rail's bid involved a proposed payment to Railtrack Group Plc of £500m, to be made on the basis of an early exit from administration. Railtrack shareholders accepted the proposal in July 2002. A further serious derailment had occurred at Potters Bar in May 2002, which was also attributed to defective points, and Railtrack had to strengthen its management of maintenance contractors before the transfer to Network Rail in October 2002, approval from the European Commission having been received in September. Whilst this went on the possibility of an interim financeability review by the rail regulator for Railtrack was brought to a halt by the secretary of state informing the regulator that if an application was made he had the power to introduce emergency legislation to prevent the review taking place. The government had intended to rationalise the existing regulatory structure by merging ORR and the SRA but this was not taken forward because the train operators wrote to the secretary of state expressing grave concerns about the threat to independent regulation, which was supported by the financiers of current franchise agreements who felt that regulation independent of government was necessary to cover non-discriminatory access, appeals and enforcement. The government proceeded with the proposal to build on the existing board structure of ORR by establishing a statutory regulatory board, hence ORR became the Office of Rail Regulation rather than the Office of the Rail Regulator.

One contradiction which arose from these developments was that whilst the government considered that the new model would require less regulation, the rail regulator considered that the loss of equity incentives meant that more regulation was required to cover areas such as business planning, management incentivisation and corporate governance. In some respects this made Network Rail a special case compared with Glas Cymru because Network Rail and the regulator did not have the benefit of the direct equity-based comparators available in the water industry, the rail regulator saying that "the regulator will require that Network Rail has regard to ORR's statements when formulating the remuneration package for its senior management team". Although independent regulation had been preserved, the rail industry remains more susceptible to government intervention and policy shifts than other utility sectors, given that strategic investment decisions are made by government in the form of the SRA.

The government carried out a rail review and published a White Paper on the Future of Rail, and the Railways Act 2005. The major changes it proposed were:

- * Transferring responsibility for safety regulation from the Health and Safety Executive to the Office of Rail Regulation, thereby making ORR a combined economic and safety regulator.

- * Changes to the price review process so that government would be required to specify the strategy for the railways as an input into price reviews.
- * The abolition of the SRA and transfer of many of its responsibilities, including franchising, direct to government and the Department for Transport.
- * Transfer of some of the SRA's responsibilities to NetworkRail which would be expected to work more in partnership with the train operators.

By 2007 the rail industry had returned to a more positive context. Passenger kilometres increased from 33bn in 1991 to 43bn in 2005; total freight moved increased by 7.5% in 2006 compared with 2005; it maintained compliance with timetable requirements which enable passengers to be given information about train journeys well in advance; between 2004 and 2005 NetworkRail achieved efficiency savings in operations, maintenance and renewals expenditure of 24%, 19% and 15% respectively; the major accident risk index had reduced from 103 in 2002 to 79 in 2006; the number of passenger complaints per 100,000 journeys decreased by 2% from 2005 to 2006, the overall passenger complaints being 72 per 100,000 journeys.

Coalition 2010 Conservative/Liberal Democrats

The new coalition returned to privatisation of state assets with the sale of Royal Mail in 2013. The shares on offer were divided between the public, institutional investors, such as city investors and pension funds, and shares reserved for Royal Mail employees. The public placed orders for more than seven times the shares available to them, which was more intense than the demand for British Gas and British Telecom at the height of the privatisation drive in the 1980s. Such oversubscription naturally raised wide criticism that the shares had been underpriced in the government's sale prospectus, which set a maximum of 330p per share. The sale was also controversial because of the public sector brand which Royal Mail and the Post Office represented with its universal service obligations to deliver in any part of the country at a common price. Even Mrs Thatcher, the pioneer of privatisation in the 1980s, had drawn the line at the Royal Mail, saying that she was not willing to have the Queen's head privatised. The controversy was also reflected in the 96% of Royal Mail's 150,000 staff who balloted to go on strike, and a proportion of them said they would not accept their free shares worth some £2,200. The government was keen to ensure that the small investors got their fair share, and shares for the institutional investors were focused on providing shares to pension funds and insurance companies that held the savings of millions of people. The government also said that the 5% of people that applied for more than £10,000 worth of shares – who were likely to be well established investors having made their applications through a stockbroker – would not receive an allocation.

The share price did rise to 600p soon after the sale in October 2013, and the criticism of the government, particularly Vince Cable, the Business secretary, increased with a report from the National Audit Office, entitled *The Privatisation of the Royal Mail*, in May 2014. It said that taxpayers lost out because the government set an overly cautious share price for the privatisation of Royal Mail. At 330p the Treasury received £2bn but the share price closing at 455p after the first day of trading, an increase of 38% in value, meant some £750m had been lost to taxpayers. Five months later the shares were worth 72% more than the offer price. The government bowed to the mounting political pressure from MPs, and Vince Cable appointed Paul Myners to review lessons learnt from the controversial sale. Lord Myners' review concentrated on Vince Cable's reliance on the advice of investment bank Lazard, which advised the government not to increase the flotation price above 330p per share, but which made an immediate profit of £8m buying and selling Royal Mail shares. Lazard's corporate advisory arm collected £1.5m for its advice that increasing the share price would put off city investors from buying shares, and Lazard's investor

division was given 'priority investor status', allowing it to buy 6m shares at 330p. Lazard insisted that there was nothing wrong with the company acting as both advisor and investor because there were 'Chinese Walls' which prevented one half of the business knowing what the other half was doing.

The sale of Royal Mail effectively completed the privatisation of utilities and network industries over the period 1984 to 2013, and all of them had a proportion of the public who believed that they should have remained in public ownership. Whilst the nationalised industries could have continued as before, there is little evidence that privatisation, competition and regulation has been inherently faulty. No political party has sought to return to nationalisation, however controversial, particularly in relation to rail, energy and mail. The system has developed over time as issues such as environmental standards or unfair competition arose, but these have been dealt with by the regulatory institutions established to deal with them. BT still has to consider whether its network provider OpenReach should be sold to avoid the accusation that it does not provide the same service to competitors as it does to BT. In the case of rail, many thought that the loss of equity incentives for NetworkRail would make it inefficient, but in fact there are many positive incentives which exist, such as, financial effects of management and employee incentive arrangements; reputational incentives from being associated with a highly successful private sector company; and motivational effects from having the opportunity to make a difference to the environment which is not subject to excessive government or regulatory interference. The fact that many other countries have adopted the basic model of privatisation and regulation developed in the UK from 1984 also supports the idea that privatisation, competition and regulation has been a success compared with monopolised nationalised industries.

7. An overview – bibliography

The experience of regulating the privatised utility and network industries over the years 1984-2011 has been an important example of political economy in practice, and particularly because it was not confined to the UK in particular, but became a global trend. Prior to the UK experiment of using 'price' control, the main approach to regulating such industries was 'rate of return' regulation (or profit control) practised in the USA. This became known as 'cost-plus' regulation, as the only incentive on suppliers was to argue for higher costs, on which they could earn a normal rate of return. The clear incentives to reduce costs under price control meant that countries in Europe, and countries such as Australia, sought to take advantage this approach. The theory of incentives and efficiency in price control has been well discussed, and the technicalities of integrating economics and accountancy have been tested at each periodic review of price controls, as have finance questions, such as what the cost of capital should be. The problem of on-going debates about how best to regulate such industries is that public confidence might be hard to find. In one sense, the underlying theory of incentive regulation is hard to test in practice, and it is probably only after some twenty or thirty years that research can now be done to find out whether incentive regulation, and the introduction of competition, has really delivered on its promises, compared with alternatives.

One debate that often comes up is whether the creation of NetworkRail, and the government taking a majority of shares in some banks to stop them going bankrupt after the financial crash in 2008, were really nationalisation, or simply a financial arrangement. In the case of both it could be said that it was public service issues which the catalyst, but at the same time the government never said that it intended to hold on to its majority share holdings in some banks, such as RBS and Lloyds, but would sell them when the time was right. It could be said that if the government wanted to change the behaviour of banks it might be useful to have a majority

shareholding so that it can influence, or require, the changes in bank behaviour that it desired. Otherwise it would have to rely on statutory changes or regulation which might not be so effective. However, the new Conservative government following the coalition has made it clear that it intends to sell their shares, and in Royal Mail too, as a contribution to reducing the deficit. One test of whether nationalisation has taken place is whether the assets and the liabilities of the company are included in the Whole of Government Accounts, something which the National Audit Office argued for in relation to NetworkRail for some years. This has now been done, so it could be said that only case which could be said to have been renationalised would be NetworkRail.

The following publications have been chosen to reflect the debate and experience. The Centre for the study of Regulated Industries (CRI), operated from 1990 to 2010, and its wide range of publications can be found at www.bath.ac.uk/cri, where the various types of publication are listed (most downloadable), including research reports, the series of regulatory reviews, lectures and occasional papers. One useful example is CRI Proceedings 31, which was a 20th anniversary collection to mark the publication of the so-called 'Littlechild Report' in 1983, just prior to the privatisation of British Telecommunications, and which established the form of regulation (reprinted as the Appendix to CRI Proceedings 31).

The following publications have been ordered by publication date, and it is interesting to see how the debate evolved over the years (and each publication has its own lists of references). In addition to these, the most important publications in terms of the practice of regulation are the annual reports and periodic price control reviews (which have many stages where there is public consultation) published by the various economic regulators of the utility and network industries. Accompanying those are reports by the Competition Commission (previously the Monopolies and Mergers Commission) where there have been appeals against the economic regulators' proposals, and by the National Audit Office (NAO).

In 1996, the NAO published its report on the work of the directors general of telecommunications, gas supply, water services and electricity supply. Its report rather validated the regulatory system, concluding that "services have been maintained and standards of service improved; effective competition has been developed in many areas; and prices have fallen in three of the four industries, and the price reviews examined tightened price control even further". The matters it raised for further consideration related to improving corporate governance, transparency and accountability. In 2002 the NAO produced a further report called Pipes and Wires, and this showed how any regulatory system has to be reviewed over time, and adapt as necessary. It concluded that "the RPI – x approach does have some inherent limitations" and linked these to "the strength of the incentives (particularly at the end of the price control period); investment in networks may be distorted; insufficient or excessive returns may be allowed to investors; and the process of price regulation may impose excessive demands on companies".

1983 Regulation of British Telecommunications' Profitability – A Report, SC Littlechild, Department of Industry

1986 Economic Regulation of Privatised Water Authorities – A Report, SC Littlechild, Department of the Environment

1991 Regulators and the Market – An Assessment of the Growth of Regulation in the UK, ed C Veljanovski, Institute of Economic Affairs

1992 Privatisation, Public Ownership and the Regulation of Natural Monopoly, CD Foster, Blackwell

- 1993** A Theory of Incentives in Procurement and Regulation, JJ Laffont and J Tirole, MIT Press
- 1994** Regulatory Reform – Economic Analysis and British Experience, M Armstrong, S Cowan and J Vickers, MIT Press
- 1995** The Regulatory Challenge, M Bishop, J Kay and C Mayer, OUP
- 1996** The Work of the Directors General of Telecommunications, Gas Supply, Water Services and Electricity Supply, National Audit Office, HC 645 Session 1995-96
- 1997** Law and the Regulators, T Prosser, OUP Clarendon Press
- 1998** A Reader on Regulation, R Baldwin, C Scott and C Hood, OUP
- 1998** Changing Regulatory Institutions in Britain and North America, eds GB Doern and S Wilks, University of Toronto Press
- 1999** Privatisation, Restructuring, and the Regulation of Network Industries, DM Newbery, MIT Press
- 1999** Understanding Regulation – Theory, Strategy and Practice, R Baldwin and M Cave, OUP
- 2001** Regulation of Network Utilities – The European Experience, eds C Henry, M Matheu and A Jeunemaitre, OUP
- 2001** HM Treasury External Efficiency Review of Utility Regulators, Final report, WS Atkins
- 2002** Pipes and Wires, National Audit Office, HC 723 Session 2001-2002
- 2003** The British Regulatory State – High Modernism and Hyper-Innovation, M Moran, OUP
- 2003** Energy, The State, and the Market – British Energy Policy since 1979, D Helm, OUP
- 2003** Competition and Regulation in Utility Markets, ed C Robinson, Institute of Economic Affairs and London Business School, Edward Elgar
- 2003** The UK Model of Utility Regulation – A 20th Anniversary collection to mark the ‘Littlechild Report’ – Retrospect and Prospect, ed I Bartle, CRI proceedings 31, School of Management, University of Bath
- 2004** Successes and Failures in Regulating and Deregulating Utilities – Evidence from the UK, Europe and the USA, ed C Robinson, Institute of Economic Affairs and London Business School, Edward Elgar
- 2007** UK Economic Regulators, House of Lords Select Committee on Regulators, HL Paper 189-I
- 2009** The Official History of Privatisation – Volume 1 The Formative Years 1970-1987, D Parker, Routledge

V The Regulatory Framework

1. Developing and Maintaining an Effective Regulatory State

The previous examples of macro and micro political economy clearly show that outcomes have to be analysed in the relevant context; the balance of interests involved and their respective powers and influence. In one sense, the practice of political economy is always about 'variance analysis'; comparing what happens with what might have happened, or should have happened. Whether the concern is at the micro or the macro level of debate, each takes place within a framework which encompasses the regulatory state. In this context, the definition of regulatory state encompasses all the ways in which the state maintains the balance and direction of activity in the country.

The state in all countries (whichever political system operates) is concerned with the elements which reflect a country and its direction:

- * The legal system and its division into criminal and civil law
- * The armed forces and the police services
- * The health and welfare services
- * Maintaining efficient markets for private sector activity
- * Balancing the rights, needs and responsibilities of different parts of society
- * Conducting foreign policy

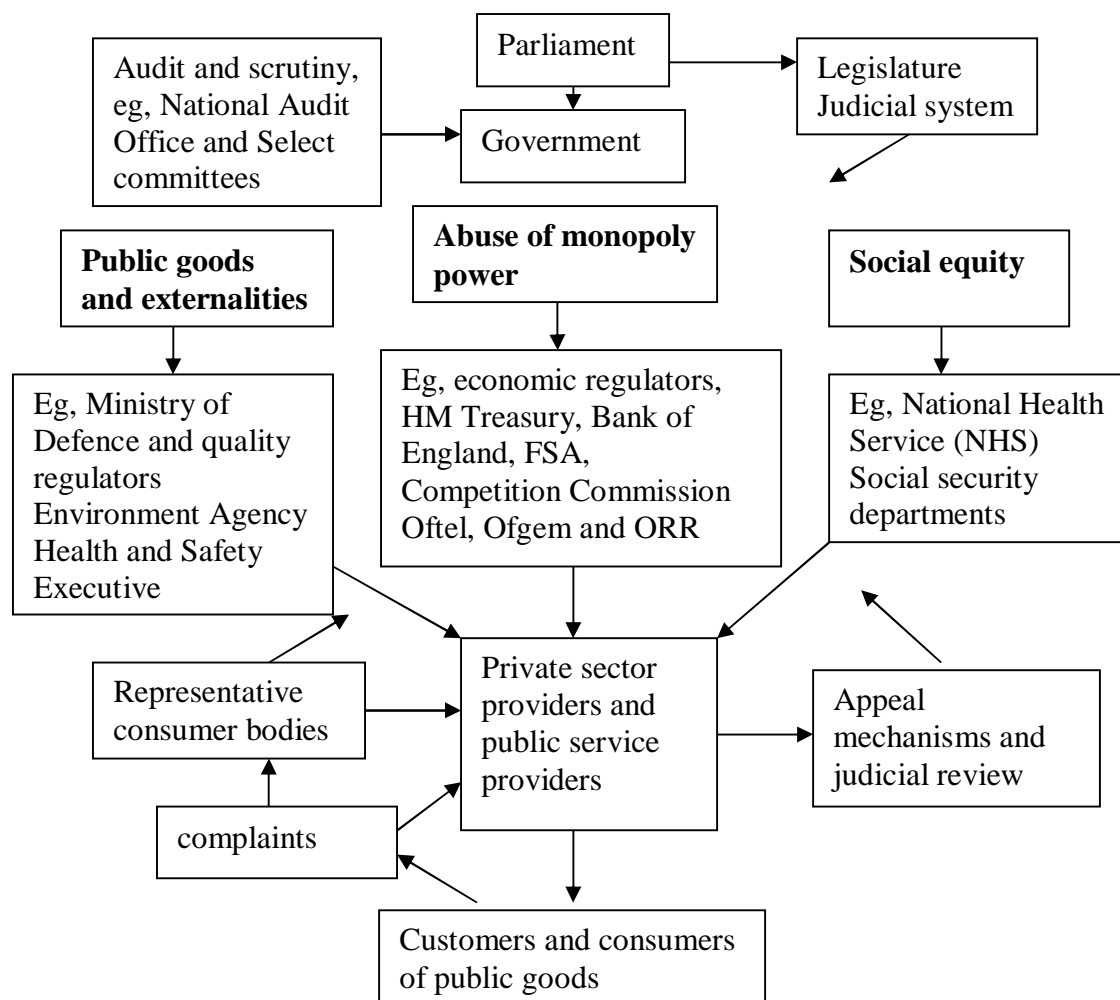
and in this context the state can be regulating by being a provider of goods and services or by regulating those who supply them.

In terms of rationalising the role of the regulatory state, it is useful to classify the various types of 'conducts' which would justify some appropriate state action. These could be represented as follows:

	<u>Analytical framework</u>	<u>Examples</u>
Non-market 'conducts'		
*Positive	public 'goods'	charitable works
*Negative	public 'bads'	racism
Market 'conducts'		
1. 'perfect' markets	inequitable outcomes	inadequate income disability
2. 'imperfect' markets	abuse of monopoly power	dominance cartels information asymmetry
	externalities	
	*positive	connection density
	*negative	polluting third parties exhausting non-renewable resources inadequate health and safety
	no private provision	defence, fire service public broadcasting

In this context it is important to see how the institutional structure of the regulatory state evolves to deal with each of these potential ‘conduct’ problems, perhaps too little of the public goods and too much of the public bads. In the UK, the framework could be illustrated as follows:

The framework of the UK regulatory state



The illustration shows how the various ‘checks and balances’ operate. The judiciary is independent of government, parliament approves (or not) the public expenditure plans of the government, and the activities of government and its departments are scrutinised by the National Audit Office or ‘Select’ Committees of the Houses of Parliament. Customers can use representative consumer bodies to take up their complaints when the provider refuses to respond appropriately, and regulated companies can appeal against decisions made by the regulators.

The key lessons of the illustrative diagram to maintaining an effective regulatory state are that:

- * There should be separation of roles and responsibilities in the regulatory state

- * Effective government requires effective accountability

and the next section concerns how this has been developed in the UK. The two lessons are not unrelated, in that creating too many separate bodies can overburden the system of communication and undermine effective accountability. A cost-effective regulatory state requires a sensible functional analysis of the roles and responsibilities to be carried out, avoiding having too many that are too small, or too few that are too large.

In practice, however, the UK regulatory framework does change as events happen, whether these are due to changes of government or the response to various ‘crises’. The overall framework may remain the same but its detail may change significantly. Some examples are worth mentioning. In 2013 the traditional audit structure will change when legislation to abolish the Audit Commission was introduced, giving wider powers to the National Audit Office, and contracting out local authority auditing to the private sector. The NAO will prepare and maintain a Code of Audit Practice, setting out the framework under which local auditors will work. One purpose of the change was to encourage cost-effectiveness, perhaps with a saving of some £1.2bn over ten years. The regional structure of the police service is also being pressurised to reflect more effectively a national network by the new Chief Inspector of Constabulary, involving more sharing of intelligence and the alignment of operational practices. Exam standards have become under increasing pressure due to repeated accusations that apparently rising standards and exam results are mostly due to grade ‘drift’. Reports by Ofsted and Ofqual may be one rational for the government seeking major changes to the national curriculum and the exam structure. The Office of National Statistics is concerned to alter the measurement of inflation by the RPI but changes can result in fierce lobbying where interests are affected, particularly by pensioners. The result is the publication of a new index, the ‘RPIJ’, based on a different way of averaging prices (the ‘Jevons’ average). Finally, a very significant change which has resulted for the 2008 global financial crisis is that because the Financial Services Authority (the FSA) was judged to have been too weak in regulation to avoid the crisis, the powers have now been separated with consumer protection being given to a new Financial Conduct Authority, and the financial prudence regulation of banks in terms of their risk and balance sheets passed to the Bank of England.

2. Better regulation principles and institutional structures

The idea of separating roles and responsibilities was well tested with the privatisation of the UK’s utilities and network industries from 1984. The economic regulators were given their own statutory duties and were responsible for the executive implementation of the overall policy set by the government, which focused on long term requirements for security of supply and environmental improvements. The economic regulators had the appropriate degree of independence and security, with the right to exercise discretion, and were expected to be free from political interference, although government could issue regulatory guidelines. The line of control was therefore that government set the policy, economic regulators ensured the delivery of the required outputs, and the regulated companies sought to deliver the outputs efficiently, with a focus on the needs of their customers. A major change of emphasis took place, however, when New Labour came to power in 1997, the focus turning to accountability of the regulators and the power of consumers.

The need for this change of emphasis might best be reflected in what Edmund Burke had to say about the reality of public life in his ‘Observations on a Publication, ‘The Present State of the Nation’ in 1769:

“People not very well grounded in the principles of public morality find a set of maxims in office ready made for them, which they assume as naturally and inevitably as any of the insignia or instruments of the situation. A certain tone of the solid and practical is immediately acquired. Every former profession of public spirit is to be considered as a debauch of youth, or, at best, as a visionary scheme of unobtainable perfection. The very idea of consistency is exploded. The convenience of the business of the day is to furnish the principle for doing it”.

By 1998 the Cabinet Office had a Better Regulation Unit, and each department of state was also required to have one. A key step was the publication of The Better Regulation Guide and

Regulatory Impact Assessment (RIAs). The foreword from the prime minister, Tony Blair, captured the direction of travel well: “I have decided that no regulatory proposal which has an impact on businesses, charities, and voluntary bodies should be considered by the government without a thorough assessment of the risks, costs and benefits, a clear analysis of who will be affected and an explanation of why non-regulatory action would be insufficient”; going on to say that “Good economic analysis will help to resolve the choices; but open public debate is the key”. The guide also “highlights the importance of communication in ensuring that we regulate in accordance with the Better Regulation Task Force’s (BRTF) five key principles of good regulation”, which it characterised as:

- * **Transparency** – be open, keep it simple, be user-friendly
- * **Accountability** – to ministers and Parliament, to the public
- * **Targeting** – regulation should focus on the problem and minimise the side effects
- * **Consistency** – be predictable, people should know where they stand
- * **Proportionality** – fit the remedy to the risk, only regulate when you need to.

The requirement for RIAs and consultation has resulted in a large increase in the amount of publicly available information, with often a sequence of consultative documents resulting from a regulator’s published response to the comments received on the initial consultation document. The RIA is a good example of cost-benefit analysis because having described the issue that gave rise to the need for regulation, it “compares various possible options for dealing with that issue. One or more non-regulatory options will normally be included. The costs and benefits of each option are identified – and quantified wherever possible – to assist informed public debate about regulation”. The section on ‘valuing non-marketed benefits and costs sets out a number of approaches, such as ‘revealed preferences’ or ‘stated preferences’. Even where such analysis is not available, an ‘implicit’ valuation can often be calculated to inform the decision maker of the meaning of their decision. Take the example of a road scheme having to demolish an historic building. One alternative might be to build a tunnel for that part of the route. If the incremental cost difference between the two routes was £2m, then if the choice was to demolish the building, then the decision is saying we value the historic building at less than £2m. Guidance on such things is available in HM Treasury’s ‘Green book’ – ‘Appraisal and Evaluation in Central Government’, and on environmental matters in the Department of the Environment’s (now DEFRA) ‘Policy Appraisal and the Environment’. One motive behind the RIAs was to reduce bureaucratic ‘red tape’, and by 2005 the BRTF published its signature document in this regard ‘Regulation – Less is More, Reducing Burdens, Improving Outcomes’. The two key conclusions were:

1. “The UK can considerably reduce the regulatory burden on business by adopting the successful Dutch approach to reducing administrative costs. This approach involves first measuring administrative burdens and then setting a target to reduce them. The golden rule is that what gets measured gets done”.
2. “We need a ‘One in, One out’ approach to new regulation, which forces departments to prioritise between new regulations and to simplify and remove existing regulations. This will complement the administrative cost reduction programme”.

This was an important publication politically, but the one in, one out rule is a difficult one to give meaning to in practice because there is no measure of scale: the one coming in might be significantly different from the one going out.

By 2006 the importance of the BRTF was recognised by giving it a more permanent footing as the Better Regulation Commission (BRC). The BRC set out what it does in this way:

- * **How we regulate** – the process of regulation and its implementation
- * **How much we regulate** – the volume of regulation and its implications
- * **Are regulations working** – specific regulations, their content and impact.

The Legislative and Regulatory Reform act was introduced in 2006 and followed on from the Hampton report in 2005 which “argued for the creation of a regulatory system, at both a national and local level, in which risk assessment was the basis for all regulators’ enforcement programmes”. The BRC felt that “Full implementation of the Hampton principles will result in fewer forms, fewer inspections, better advice, and better co-ordination between regulatory bodies”. As the UK is a member of the European Union, the Cabinet Office sought to ensure RIAs were applied to European legislation and directives, publishing in 1999 its ‘Guide to Better European Regulation’. By 2006 the European Council requested the European Commission to publish an ambitious ‘Communication’ on better regulation which asked the Spring European Council in March 2007 to endorse a 25% target to reduce administrative burdens by 2012 in nine priority areas of both EU and national legislation (being company law, health protection, employment law, fiscal law/VAT, statistics, agriculture, food labelling, transport and fisheries).

Missing principles and institutional change

The progress in terms of developing and maintaining an efficient regulatory state was important in the years 1980 to 2010, but even so there remains much to discuss. First, there is the five principles of better regulation. These play an important role, but do they cover the full picture, and are there missing principles? The Department of Business Innovation and Skills introduced new principles in 2011, but these only appear to deal with one, Coherence, the other new ones being variants of the existing ones. They were: efficiency, focus, predictability, coherence, adaptability and accountability. If the functions to which the principles should apply were divided into three parts: Policy principles; Implementation principles and Governance principles, then it seems that four missing principles could usefully be added (objectivity, coherence, completeness, and credible commitment) as shown in the following:

Functions

Policy Principles

Principles

- *objectivity
- *coherence
- *completeness
- *credible commitment

Implementation Principles

- *proportionality
- *targeting
- *consistency

Governance Principles

- *transparency
- *accountability

In applying these principles, the policy function should have regard to public risk analysis, the polluter pays principle, the precautionary principle, and the necessary effort to demonstrate credible commitment over the longer term. This was an important issue for economic regulators and the privatised utility and network industries in terms of the cost of capital because investors could not be sure that with five year price reviews, the long term investments would be

treated fairly at each review. The implementation function should have regard to the cost-benefit test, the integration of quantitative and qualitative information, the proper use of incremental costs and benefits in setting standards, and specifying correctly the range of do-nothing and do-something options. The principle of consistency is always an interesting one for debate, given sometimes there is a choice of being consistent with past decisions or being consistent with the underlying purpose of the regulation. The governance function should have regard to regulators giving reasons for decisions, and for effective scrutiny to include the possibility of independent review.

For good government, the UK government's five principles of sustainable development make good complements to the principles of better regulation in terms of political economy:

- * A sound economy
- * Living within environmental limits
- * A fair and just society
- * Using sound science wisely
- * Promoting good governance

However, to make these felt it is important to ensure that the principles are not only codified, but institutionalised in terms of regulatory processes and the appropriate separation of roles and responsibilities.

Second, the standing of the Director General as 'the' regulator. The model adopted at privatisation in the UK was to give the regulatory powers and duties to an appointed individual. However, by 2000, there was a criticism of this model in terms of it being 'over-personalised'. It was thought better to make a board responsible, on the traditional company model, with a chairman and a chief executive. The Utilities Act 2000 took the first step by replacing the Director General of electricity and gas supply with GEMA, the Gas and Electricity Markets Authority. Other regulated industries followed, but there is one problem with the new model, which is effective accountability. The Director General was very exposed to questioning and scrutiny (certainly on television), and there is a strong incentive on the regulator to do an effective job. This is less so with a board structure, and the boards were in effect obliged to nominate one person to represent the 'public face' of the regulator. It is interesting that the term GEMA is rarely seen in relation to energy regulation, the task falling to the chief executive of Ofgem, where the chief executive effectively substitutes for the previous director general.

Thirdly, the place of consumer representative bodies. The Utilities Act 2000 introduced Energywatch as the 'independent consumer watchdog, taking over from the previous Gas Consumers Council (which had been separate from Ofgas) and the regional electricity consumer committees (which were part of Offer). This development might seem curious given the Utilities Act also formalised the primary duty of the utility regulators to be 'consumer protection'. The artificiality of it was soon shown by the regulators and the consumer bodies issuing 'memorandums of understanding'. However, it should be recognised that while one regulatory duty was to protect consumers through price controls, another duty was to ensure that investors could earn a return at least equal to the cost of capital, given the risks involved. In terms of balance, it makes sense for the regulator to sit between the shareholder and the customer in achieving an efficient and equitable balance.

The political economy of the regulatory state, as these examples have shown, is one of development, but one concern could be that the names and roles of departments change so regularly as political parties come into power and others leave, that it is difficult for the general public to become knowledgeable about them and to feel more engaged with the process. The

National Consumer Council changed its name to Consumer Focus, Ombudsmen were introduced to adjudicate on various sectors, and the Department of the Environment (DoE) lost one of its main roles to the new Department for Energy and Climate Change (DECC), being renamed the Department of the Environment, Food and Rural Affairs (DEFRA). The other concern is that the public in general are resistant to a balanced view. As academic Peter Jackson said in an article in 2011 ‘Governance by Numbers: what have we learned over the past 30 years’, “The performance of public sector organisations, even when they are doing well, is the subject of critical commentary. No matter what they do, some group will argue that they could do better and do things differently. Criticism of public service organisations is a very public affair. That is the nature of the public sector. The management of private sector organisations are seldom placed in the public gaze unless they are involved in some major environmental catastrophe or a governance scandal”.

VI Appendix

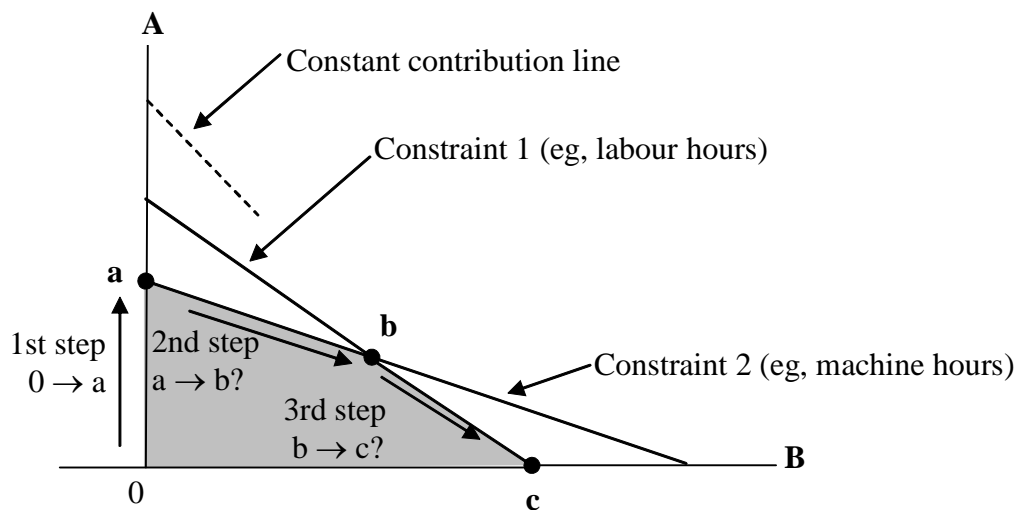
Linear programming

– a ‘search’ technique for the optimal solution

The following illustrates, first, the ‘path’ of a search technique (using two examples), which is then followed by worked examples of the ‘Simplex’ method used in computer programmes. It shows how, at each point on the path, by substitution, the next choice of path is based on comparing the gain from moving on with the position currently arrived at (so that figures are converted to be relative to the current benchmark point on the path under consideration).

Example 1: maximise the contribution (ie, price less variable cost) from two products (A and B); two production constraints (eg, total machine hours and labour hours available). A has the highest contribution (or NPV) per unit.

The ‘path’ from 0 to the optimum (point a, b or c)



1st step: given A has the highest contribution per unit (or NPV), produce maximum possible of A (ie, at point a). If B had had the highest contribution (ie, the slope of the constant contribution line was greater than 45°), then the ‘path’ would have gone in the other direction, that is $0 \rightarrow c$, then consider $c \rightarrow b$ etc.

2nd step: given you are at a, is it worth moving to b? Note constraint 1 is a ‘non-binding’ constraint at a, so that the only opportunity cost to producing B is having to produce less of A to release units of the binding scarce resource (in this case machine hours - constraint 2). Calculations of the ‘net’ gain from substituting units of B for A will tell you whether moving from point $a \rightarrow b$ is worthwhile.

3rd step: if it was worth moving to point b, then is it worth moving to point c? The same logic as at the second step applies.

Note: point a is optimum if A has ‘higher ratios of contribution to constraint for both production constraints’. Point c is optimum if B has both ratios higher, and point b is optimum where A is higher for one and B is higher for the other. This is determined by the slope of the constant contribution (or NPV) line, relative to the slopes of the production constraint lines.

The following figures would have resulted in all three steps shown in the above diagram, moving through points a and b to point c, the optimum decision to produce only B. This is because B has higher contribution to constraint ratios for both constraints, and the path taken is because A has the highest contribution.

Example 2: an illustration with 3 products (A, B and C) and one constraint, showing the branching choice at point a and how the constraint ‘line’ becomes a constraint ‘plane’ with three products.

2nd step: the calculation of whether product B or C has the highest ‘net’ contribution, allowing for the off-setting reductions in A, will determine whether you move to point b or c, or stay at a.

A worked example of this simple three product, one constraint case is given at the end, which follows on the application of the Simplex method in practice.

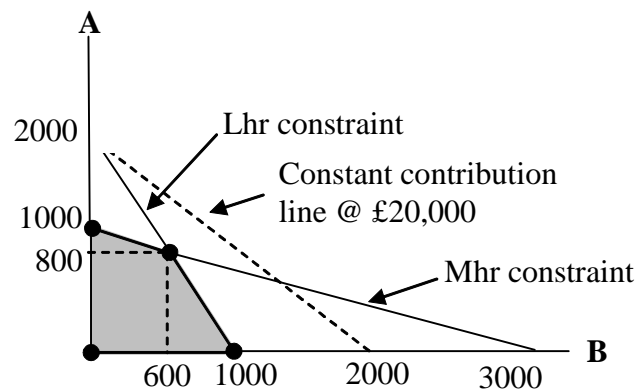
I Solving a linear programme – the simplex method

Primal programme

Maximise $10A + 10B$
Subject to:
 Mhr constraint $3A + 1B \leq 3000$
 Lhr constraint $1A + 2B \leq 2000$

Note – This formulation ignores the **non-negativity constraints** that are also needed to fully define the production possibility area, ie, $A; B \geq 0$

Primal diagram



Step 1: Introduce 'slack' variables (S) to make the constraints equalities (so that the unused resources from a 'non-binding' constraint are shown in the slack variable), ie:

$$\begin{aligned} 3A + 1B + S_{Mh} &= 3000 \\ 1A + 2B + S_{Lh} &= 2000 \end{aligned}$$

Step 2: Rearrange the constraints so that the 'slack' variables are on the right hand side, ie,

Objective (max)	10A	+ 10B	= C	Contribution
Mhr constraint	- 3A	- 1B	+ 3000	= S_{Mh}
Lhr constraint	- 1A	- 2B	+ 2000	= S_{Lh}
ie,	product A	product B	quantities available	slack variables

Step 3: Form the '**first**' matrix where production of A and B is zero, ie, where $A = 0$ and $B = 0$, then:

the 'first' matrix

Objective	10A	+ 10B	= 0	total contribution (C)
Mhr constraint	-3A	- 1B	+ 3000	= 3000
Lhr constraint	-1A	- 2B	+ 2000	= 2000

Step 4: Choose the product with the highest contribution (in this case both products are equal, so choose A) to produce the '**second**' matrix.

Step 5: Find the constraint which limits the maximum production of A (clearly machine hrs from the diagram above), ie:

$$\begin{aligned} \text{Mhrs} \quad 3000 \div 3 &= 1000 & \therefore 1000A \text{ is the maximum possible} \\ \text{Lhrs} \quad 2000 \div 1 &= 2000 \end{aligned}$$

Step 6: Rearrange the relevant constraint line which is now 'binding' with this production level (ie, Mhrs) to put A in terms of the other variables, ie:

$$\begin{aligned} \text{Mhr constraint} \quad &- 3A - 1B + 3000 = S_{Mh} \\ \text{therefore} \quad &A = - \frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh} \end{aligned}$$

Step 7: Substitute this value of A (in terms of the other variables) into the **other** equation lines of the ‘first’ matrix, ie:

$$\begin{aligned}\text{Objective-max} & 10(-\frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh}) + 10B = C \\ \text{Lhrs constraint} & -1(-\frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh}) - 2B + 2000 = S_{Lh}\end{aligned}$$

Step 8: Simplify and incorporate with the rearranged Mhr constraint (with A in terms of the other variables) to form the **‘second’ matrix**, ie, given:

$$\begin{aligned}\text{Objective} & -3\frac{1}{3}B + 10,000 - 3\frac{1}{3}S_{Mh} + 10B = C \\ \Rightarrow & +6\frac{2}{3}B - 3\frac{1}{3}S_{Mh} + 10,000 = C \\ \text{Lhrs constraint} & \frac{1}{3}B + \frac{1}{3}S_{Mh} - 1000 - 2B + 2000 = S_{Lh} \\ \Rightarrow & -1\frac{2}{3}B + \frac{1}{3}S_{Mh} + 1000 = S_{Lh}\end{aligned}$$

the ‘second’ matrix

Objective	+ 6$\frac{2}{3}$B	- 3$\frac{1}{3}$S_{Mh}	+ 10,000	= C
Product A constraint	- $\frac{1}{3}$B	- $\frac{1}{3}$S_{Mh}	+ 1000	= A
Lhr constraint	- 1$\frac{2}{3}$B	+ $\frac{1}{3}$ S_{Mh}	+ 1000	= S_{Lh}

and note that S_{Mh} and B are zero so that the quantities column gives the results for contribution (C), the production of A, and the amount of labour hours left unused (S_{Lh})

Step 9: The optimum solution is achieved when the contribution row (objective) contains only negative or zero values (other than the positive contribution of course)

Product B shows that it would generate a positive contribution from being produced. Hence we go through the process again to form a **‘third’ matrix**, which we know from this simplified example will be the final, optimum solution.

The net contribution of 6 $\frac{2}{3}$ from producing B arises because we are now **substituting B** for units of A, given the starting point of currently producing 1000 units of A. This can be explained as follows (given labour hours are unconstrained at present, and so to produce a unit of B using machine hours requires a reduction in A, given all machine hours have been utilised producing 1000 A):

- 1 unit of B requires 1Mhr, therefore $\frac{1}{3}$ of a unit less of A must be produced to release the 1 Mhr. A unit of A yields a contribution of 10 and therefore the total contribution from producing A goes down by 3 $\frac{1}{3}$; to be offset by a contribution of 10 from producing a unit of B. The net gain is 6 $\frac{2}{3}$ from producing B, as shown in the second matrix.

Note that the substitution of B for A will have released $\frac{1}{3}$ Lhr from the reduced production of A, to be offset by the use of 2 Lhrs to produce a unit of B. The net use of the unconstrained labour resource is therefore 1 $\frac{2}{3}$ Lhrs per unit of B. **All of the above net figures can be identified in the ‘second’ matrix!**

Step 10: As before, we divide the constraint quantities available by the inputs required per unit of B, noting that we are only concerned with the negative values (ie, where resources are used from producing B, rather than positive values, where resources are released from producing B in substitution for something else). Hence, from the second matrix, the labour constraint is the binding constraint:

$$\begin{aligned}\text{re: A} & 1000 \div -\frac{1}{3} = -3000 \\ \text{re: Lhr constraint} & 1000 \div -1\frac{2}{3} = -600\end{aligned}$$

Step 11: Rearrange the now binding labour hour constraint line as before so that B is in terms of the other variables, ie:

$$-1\frac{2}{3}B + \frac{1}{3}S_{Mh} + 1000 = S_{Lh}$$

$$\Rightarrow B = -0.6S_{Lh} + 0.2S_{Mh} + 600$$

Step 12: Substitute for B in the other equations of the second matrix to give the ‘third’ matrix as follows:

$$\begin{aligned} \text{Objective} \quad & 6\frac{2}{3}(-0.6S_{Lh} + 0.2S_{Mh} + 600) - 3\frac{1}{3}S_{Mh} + 10,000 = C \\ \Rightarrow \quad & -4S_{Lh} + 1\frac{1}{3}S_{Mh} + 4000 - 3\frac{1}{3}S_{Mh} + 10,000 = C \\ \Rightarrow \quad & -4S_{Lh} - 2S_{Mh} + 14000 = C \\ \\ \text{Product A} \quad & -\frac{1}{3}(-0.6S_{Lh} + 0.2S_{Mh} + 600) - \frac{1}{3}S_{Mh} + 1000 = A \\ \Rightarrow \quad & +0.2S_{Lh} - 0.06667S_{Mh} - 200 - \frac{1}{3}S_{Mh} + 1000 = A \\ \Rightarrow \quad & +0.2S_{Lh} - 0.4S_{Mh} + 800 = A \end{aligned}$$

the ‘third’ matrix

Objective	- 4S_{Lh}	- 2S_{Mh}	+ 14,000	= C
Product A	+ 0.2S_{Lh}	- 0.4S_{Mh}	+ 800	= A
Product B	- 0.6S_{Lh}	+ 0.2S_{Mh}	+ 600	= B

Step 13: Which is the right answer! A = 800; B = 600; Contribution = 14000

Note that the ‘third’ matrix has only negative numbers in the contribution row (objective), therefore no further gains can be made. **Note** that S_{Lh} and S_{Mh} are zero because both constraints are binding. **Note too** that the figures in the ‘objective’ line for the final matrix (slack variables S_{Lh} and S_{Mh}) are the ‘**internal**’ **opportunity costs** of labour and machine hours (ie, 4 and 2 respectively). **More generally**, the ‘final’ matrix would have a row of binding constraint ‘slack’ variables, which by definition are zero, accompanied by products ‘excluded’ from the optimal plan, and a column of products which are in the production plan, plus slack variables for non-binding constraints.

For example:		Products not in the optimal plan $M, N, \dots = 0$	Binding constraints $S_1 \quad S_2 \quad S_3 \dots = 0$	Quantities $X = \text{Contribution}$
products included in the optimal plan	A		The number of binding constraints normally equals the number of products in the optimal plan	X Those in the production plan,
	B			X therefore $X > 0$
	C			X (Those products excluded,
	:			: then $M, N, \dots = 0$)
non-binding constraints	S_4			X eg, non-negativity constraints,
	S_5			X where the slack variable value
	S_6			X will reflect the production
	:			: decisions above

Note that the coefficients on the products ‘excluded’ from the optimal plan in the contribution (ie, objective) row will show the opportunity cost of wrongly including a unit of an excluded product into the production plan (see worked example II which follows).

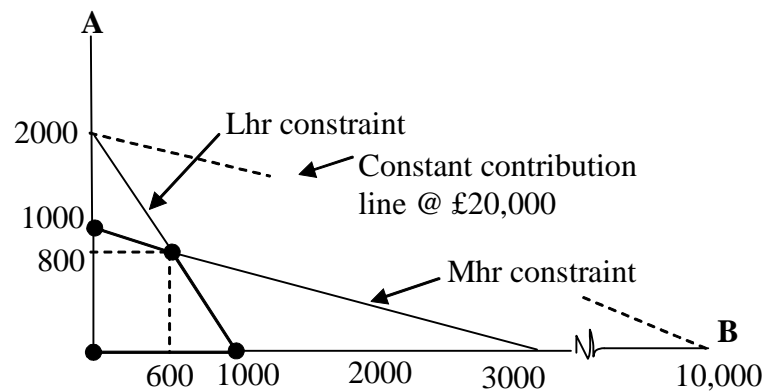
II Solving a linear programme – continued

Adapting the example to demonstrate the general ‘final’ matrix, including non-negativity constraints (the ‘second’ matrix becomes the final matrix)

Primal programme

Maximise $10A + 2B$
Subject to:
 Mhr constraint $3A + 1B \leq 3000$
 Lhr constraint $1A + 2B \leq 2000$
 Non-negativity constraints $A \geq 0$
 $B \geq 0$

Primal diagram



Step 1: Introduce ‘slack’ variables (S) to make the constraints equalities (so that the unused resources from a ‘non-binding’ constraint are shown in the slack variable), ie:

$$\begin{aligned} 3A + 1B + S_{Mh} &= 3000 & A - S_{nn}^A &= 0 \\ 1A + 2B + S_{Lh} &= 2000 & B - S_{nn}^B &= 0 \end{aligned}$$

Step 2: Rearrange the constraints so that the ‘slack’ variables are on the right hand side, ie,

Objective	10A	+ 2B	= C	Contribution
Mhr constraint	- 3A	- 1B	+ 3000	= S_{Mh}
Lhr constraint	- 1A	- 2B	+ 2000	= S_{Lh}
Non-negativity A	A			= S_{nn}^A
Non-negativity B		B		= S_{nn}^B
ie,	product A	product B	quantities available	slack variables

Step 3: Form the ‘first’ matrix where production of A and B is zero, ie, where $A = 0$ and $B = 0$, then:

the ‘first’ matrix

Objective	10A	+ 2B	= 0	total contribution (C)
Mhr constraint	-3A	- 1B	+ 3000	= 3000
Lhr constraint	-1A	- 2B	+ 2000	= 2000
Non-negativity A	A			= 0
Non-negativity B		B		= 0

Step 4: Choose the product with the highest contribution (A) to produce the ‘second’ matrix.

Step 5: Find the constraint which limits the maximum production of A (clearly machine hrs from the diagram above), ie:

$$\begin{array}{lcl} \text{Mhrs} & 3000 \div 3 & = 1000 \\ \text{Lhrs} & 2000 \div 1 & = 2000 \end{array} \quad \therefore 1000A \text{ is the maximum possible}$$

Step 6: Rearrange the relevant constraint line which is now ‘binding’ with this production level (ie, Mhrs) to put A in terms of the other variables, ie:

$$\begin{array}{l} \text{Mhr constraint} \quad - 3A - 1B + 3000 = S_{Mh} \\ \text{therefore} \quad A = -\frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh} \end{array}$$

Step 7: Substitute this value of A (in terms of the other variables) into the **other** equation lines of the ‘first’ matrix, ie:

$$\begin{array}{lcl} \text{Objective} & 10(-\frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh}) + 2B & = C \\ \text{Lhrs constraint} & -1(-\frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh}) - 2B + 2000 & = S_{Lh} \\ \text{Non-negativity A} & -\frac{1}{3}B + 1000 - \frac{1}{3}S_{Mh} & = S_{nn}^A \end{array}$$

Step 8: Simplify and incorporate with the rearranged Mhr constraint (with A in terms of the other variables) to form the ‘**second**’ matrix, ie, given:

$$\begin{array}{lcl} \text{Objective} & -3\frac{1}{3}B + 10,000 - 3\frac{1}{3}S_{Mh} + 2B & = C \\ \Rightarrow & -1\frac{1}{3}B - 3\frac{1}{3}S_{Mh} + 10,000 & = C \\ \text{Lhrs constraint} & \frac{1}{3}B + \frac{1}{3}S_{Mh} - 1000 - 2B + 2000 & = S_{Lh} \\ \Rightarrow & -1\frac{2}{3}B + \frac{1}{3}S_{Mh} + 1000 & = S_{Lh} \end{array}$$

the ‘second’ matrix

Objective	- 1 $\frac{1}{3}$ B	- 3 $\frac{1}{3}$ S _{Mh}	+ 10,000	= C
Product A constraint	- $\frac{1}{3}$ B	- $\frac{1}{3}$ S _{Mh}	+ 1000	= A
Lhr constraint	- 1 $\frac{2}{3}$ B	+ $\frac{1}{3}$ S _{Mh}	+ 1000	= S _{Lh}
Non-negativity A	- $\frac{1}{3}$ B	- $\frac{1}{3}$ S _{Mh}	+ 1000	= S _{nn} ^A
Non-negativity B	B			= S _{nn} ^B

and note that S_{Mh} and B are zero so that the quantities column gives the results for contribution (C), the production of A, and the amount of labour hours left unused (S_{Lh}). The coefficients in the objective row show the opportunity costs of the machine hour constraint and of B, if B was wrongly introduced into the optimal programme. **Note** that the opportunity cost of labour hours is zero and that the machine hours ‘internal’ opportunity cost times the total machine hours available equals the total contribution (ie, 3 $\frac{1}{3}$ x 3000 Mh = 10,000).

Step 9: The optimum solution is achieved when the contribution row (objective) contains only negative or zero values (other than the positive contribution of course)

The ‘second’ matrix is therefore the final matrix with:

$$\begin{array}{l} A = 1000; B = 0; \text{contribution} = 10,000; \\ S_{Mh} = 0; S_{Lh} = 1000; S_{nn}^A = 1000; S_{nn}^B = 0. \end{array}$$

Note that there is one product (A) in the optimal plan with one binding constraint, S_{Mh}, and one product excluded (B) with one binding constraint S_{nn}^B.

Hence there are as many binding constraints as there are potential products in the production plan.

III Solving a linear programme – continued

The three product, one constraint worked example to illustrate the ‘branching’ decision and the effect on the series of matrices

Given the following figures and the primary programme, the Simplex steps lead to three matrices.

	A	B	C	Total	ratios		
Contribution	10	8	6		A	B	C
Constraint 1	5	2	2	100	2	4	3

Primal programme

Maximise $10A + 8B + 6C$

Subject to:

Constraint 1 $5A + 2B + 2C \leq 100$

The first matrix

Objective	$10A + 8B + 6C = \text{Contribution}$
Constraint	$-5A - 2B - 2C + 100 = \text{Slack variable(s)}$

Step 1: Given A has the highest contribution, then the maximum production of A is given by $100 \div 5 = 20A$ (which would yield a total contribution of £200).

Step 2: Rearrange the constraint line to put A in terms of the other variables:

$$-5A - 2B - 2C + 100 = S$$

$$\therefore -5A = 2B + 2C - 100 + S$$

$$A = -\frac{2}{5}B - \frac{2}{5}C + 20 - \frac{1}{5}S$$

Step 3: Substitute this into the ‘objective’ line, so that:

$$10(-\frac{2}{5}B - \frac{2}{5}C + 20 - \frac{1}{5}S) + 8B + 6C = \text{contribution}$$

$$\therefore -4B - 4C + 200 - 2S + 8B + 6C = \text{contribution}$$

The second matrix

Objective	$+4B + 2C - 2S + 200 = \text{Contribution}$
Constraint (A)	$-\frac{2}{5}B - \frac{2}{5}C - \frac{1}{5}S + 20 = A$

The objective line still has ‘net’ positive values for both product B and C, but B is the highest, so choose B.

Step 4: The constraint on producing B is now A, and the maximum B is therefore given by

$$20 \div \frac{2}{5} = 50 \text{ (because to produce one unit of B requires a } \frac{2}{5} \text{ of a unit reduction in A).}$$

Step 5: Rearrange the constraint line to put B in terms of the other variables:

$$-\frac{2}{5}B = \frac{2}{5}C + \frac{1}{5}S - 20 + A$$

$$\therefore B = -1C - 0.5S + 50 - 2.5A$$

Step 6: Substitute this into the objective line, so that:

$$4(-1C - 0.5S + 50 - 2.5A) + 2C - 2S + 200 = \text{contribution}$$

$$\therefore -4C - 2S + 200 - 10A + 2C - 2S + 200 = \text{contribution}$$

The third and final matrix

Objective	$-10A - 2C - 4S + 400 =$	Contribution
Constraint (B)	$-2.5A - 1C - 0.5S + 50 =$	B

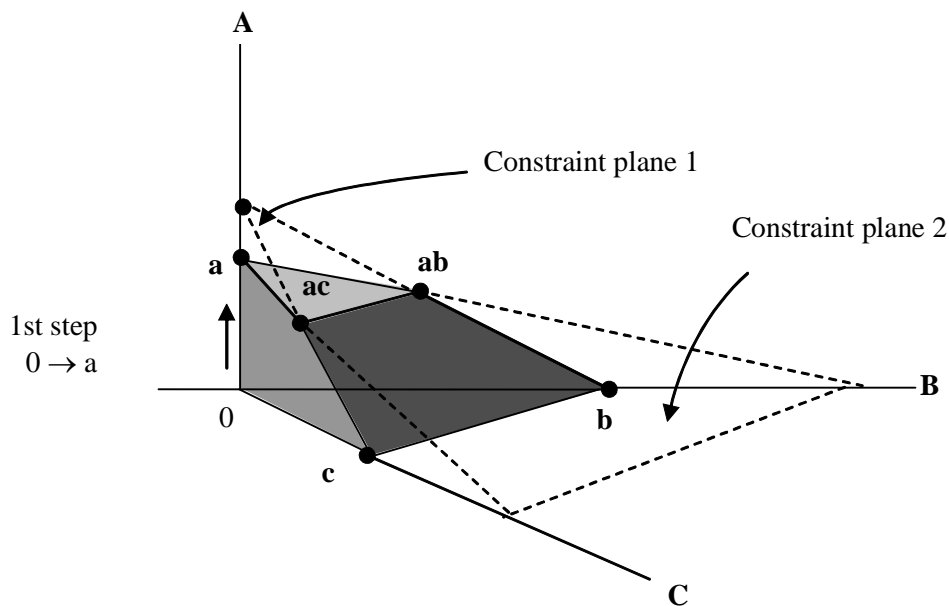
Step 7: The objective row now only has negative values (other than the maximum contribution of £400) and so represents the optimum solution, produce only B, with a maximum of 50 units, yielding a total contribution of £400 (ie, 50×8).

Note: this worked example shows how products can come into, and out of, the production plan depending on the path taken in solving the Simplex programme. In this example the 'objective' row started with A, B and C, then moved to B, C and S and then to A, C and S; remembering that when in the objective row, the value of these variables is zero!

In passing:

Given the simple illustration of a single 'plane' constraint with three products, it is worth showing that if we had two constraints then we could have 'intersecting' planes, leading to the possibility of 'mixed' production decisions (ie, a combination of ab; ac or bc with two constraints or abc with three constraints). The following shows the 'plane' constraint diagram extended to two constraints.

The 'feasible' production area with three products/two constraints



Finally:

In the worked example we noted how products left, and came back into, the objective line as we progressed along the search paths. In the first example given (two products

with two constraints), illustrative figures were given which would have given a path from a, through b to c. This means that at point a, product A has left the objective line; at point b, both products A and B have left (because both are now in the production plan) but at point c, A has been re-introduced into the objective line, substituting for the slack variable/constraint which becomes non-binding at point c. The following diagram illustrates why at point b, the matrix would have had a slack variable with a 'positive' coefficient, which means that it would be worth moving to c (whereas it would have been a negative coefficient at point a).

Why the slack variable has a 'positive' coefficient at b, hence move to c.

